

University of Groningen

Bayesian cosmic density field inference from redshift space dark matter maps

Bos, E. G. Patrick; Kitaura, Francisco-Shu; van de Weygaert, Rien

Published in:
Monthly Notices of the Royal Astronomical Society

DOI:
[10.1093/mnras/stz1864](https://doi.org/10.1093/mnras/stz1864)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2019

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Bos, E. G. P., Kitaura, F.-S., & van de Weygaert, R. (2019). Bayesian cosmic density field inference from redshift space dark matter maps. *Monthly Notices of the Royal Astronomical Society*, 488, 2573-2604. <https://doi.org/10.1093/mnras/stz1864>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Bayesian cosmic density field inference from redshift space dark matter maps

E. G. Patrick Bos^{1,2★}, Francisco-Shu Kitaura^{3,4} and Rien van de Weygaert²

¹*Netherlands eScience Center, Science Park 140, 1098XG Amsterdam, the Netherlands*

²*Kapteyn Astronomical Institute, University of Groningen, PO box 800, 9700AV Groningen, the Netherlands*

³*Instituto de Astrofísica de Canarias (IAC), Calle Via Lactea s/n, 38200 La Laguna, Tenerife, Spain*

⁴*Departamento de Astrofísica, Universidad de La Laguna (ULL), E-38206 La Laguna, Tenerife, Spain*

Accepted 2019 June 24. Received 2019 June 8; in original form 2018 October 11

ABSTRACT

We present a self-consistent Bayesian formalism to sample the primordial density fields compatible with a set of dark matter density tracers after a cosmic evolution observed in redshift space. Previous works on density reconstruction did not self-consistently consider redshift space distortions or included an additional iterative distortion correction step. We present here the analytic solution of coherent flows within a Hamiltonian Monte Carlo posterior sampling of the primordial density field. We test our method within the Zel'dovich approximation, presenting also an analytic solution including tidal fields and spherical collapse on small scales. Our resulting reconstructed fields are isotropic and their power spectra are unbiased compared to the true field defined by our mock observations. Novel algorithmic implementations are introduced regarding the mass assignment kernels when defining the dark matter density field and optimization of the time-step in the Hamiltonian equations of motions. Our algorithm, dubbed BARCODE, promises to be specially suited for analysis of the dark matter cosmic web down to scales of a few megaparsecs. This large-scale structure is implied by the observed spatial distribution of galaxy clusters – such as obtained from X-ray, Sunyaev–Zel'dovich, or weak lensing surveys – as well as that of the intergalactic medium sampled by the Ly α forest or perhaps even by deep hydrogen intensity mapping. In these cases, virialized motions are negligible, and the tracers cannot be modelled as point-like objects. It could be used in all of these contexts as a baryon acoustic oscillation reconstruction algorithm.

Key words: methods: analytical – methods: statistical – galaxies: distances and redshifts – cosmology: observations – large-scale structure of Universe.

1 INTRODUCTION

The cosmic web of the Universe arises from the gravitational instability caused by tiny primordial density perturbations, which presumably have their origin in quantum fluctuations. At initial cosmic times, perturbations are linear and the statistics describing them is extremely close to Gaussian, though some deviations from Gaussianity could be expected depending on the inflationary phase the Universe has probably experienced after its birth (Starobinsky 1980; Planck Collaboration X 2018a). Therefore, the cosmic web encodes the information to understand non-linear structure formation (Bond, Kofman & Pogosyan 1996; van de Weygaert 1996; Alpaslan et al. 2014; Cautun et al. 2014) and disentangle the interplay between dark matter and dark energy (Park & Lee 2007; Bos et al. 2012; Vogelsberger et al. 2014).

One of the great challenges of the coming years is to rigorously test these theories using observations (Werner et al. 2008; Lee et al. 2016; Connor et al. 2018; Tanimura et al. 2018). However, observations are based on biased tracers, which are affected by their proper motions, as they are measured in so-called redshift space (RSS; Jackson 1972; Sargent & Turner 1977; Kaiser 1987; Hamilton 1998). To test our theories, we need to make the step from these biased tracers to the full underlying matter density field that they probe. This requires us to make completely explicit the entire process by which the biased observations were generated from the physical world that our theories stipulate. In this, we face three main challenges: one is the action of gravity linking the linear primordial fluctuations to the final non-linear density field, the second one is modelling the bias of the dark matter tracers, and the third one is solving for redshift space distortions (RSDs). In this paper, we present a novel approach to the latter problem in the context of Bayesian reconstruction of primordial density fields given biased observations of clusters of galaxies at $z = 0$.

★ E-mail: egpbos@gmail.com

1.1 Reconstruction of primordial density fields

The topic of reconstruction of primordial density fields underwent great progress in the late 80s and the early 90s. Peebles (1989) first proposed a least action method linking the trajectories of particles from initial to final cosmic times. Weinberg (1992) proposed a rank-ordering scheme, realizing the different statistical nature between the initial and final density fluctuations. Nusser & Dekel (1992) proposed a time reversal machine based on the Zel'dovich approximation (ZA; Zel'dovich 1970). Since these pioneering works, however, there have recently been dramatic developments. The minimization of an action led to two approaches: the FAM method (Nusser & Branchini 2000; Branchini, Eldar & Nusser 2002) and the MAK method (Brenier et al. 2003; Mohayaee et al. 2006; Lavaux & Wandelt 2010). Time reversal machines based on higher order corrections have been proposed (Gramann 1993; Monaco & Efsthathiou 1999; Kitaura & Angulo 2012). The same concept has been proposed to enhance signatures encoded in the primordial density fluctuations, such as the baryon acoustic oscillations (BAOs; see Eisenstein et al. 2007), as this linearizes the density field transferring information from the higher order statistics to the two-point correlation function (see Schmittfull et al. 2015). Constrained Gaussian random field simulations (Bertschinger 1987; Hoffman & Ribak 1991; van de Weygaert & Bertschinger 1996) have been applied to study the Local Universe (Gottloeber, Hoffman & Yepes 2010), but it is hard to find a configuration of constrained peaks that give rise to a desired cosmic web configuration after a non-linear gravitational evolution. The reverse ZA can provide a first-order correction to this limitation of constrained simulations (Doumler et al. 2013; Sorce & Tempel 2018). The problem of these methods is that they are limited by the approximation of reversing gravity, which breaks down with so-called shell crossing (Hidding, van de Weygaert & Shandarin 2016).

Bayesian approaches have permitted the incorporation of forward modelling, which solves this latter problem, at a considerably higher computational cost (Jasche & Wandelt 2013; Kitaura 2013; Wang et al. 2013). We will follow here this approach and more specifically the setting described in Wang et al. (2013) based on the combination of a Gaussian prior describing the statistical nature of the primordial fluctuations and a Gaussian likelihood minimizing the squared distance between the reconstructed density and the true one. None the less, we would like to remark that the algorithmic approach of connecting the primordial density field with the final one followed by Wang et al. (2013) is based upon the one described in Jasche & Wandelt (2013), as we also do here.

1.2 Self-consistent redshift space treatment

Our novel contribution to the Bayesian reconstruction formalism consists of including coherent RSD in an analytical way, within the Hamiltonian Monte Carlo sampling of the posterior distribution function (resulting from the product of the prior and the likelihood). Nusser & Davis (1994) and Fisher et al. (1995) have also taken the approach of a self-consistent formalism for reconstructing real space densities from the biased RSS tracers that are galaxies. Both derive their formalism in linear theory, which allows Nusser & Davis (1994) to use the ZA to correct for RSDs, while for Fisher et al. (1995) it allows an analytical one-to-one transformation of the radial harmonics of the field from redshift to real space via a matrix inversion. While Jasche & Wandelt (2013) do not give a treatment of RSD, Wang et al. (2013) rely on an iterative RSD correction applied prior to the Bayesian reconstruction method (Wang et al. 2009,

2012). In a recent study in the context of Bayesian reconstructions of the primordial fluctuations, Jasche & Lavaux (2018) apply a RSS transformation to the dark matter field, in which the bias is defined with respect to RSS. Our formalism entails a more natural approach, and includes a biasing scheme with respect to Eulerian space (see also Bos 2016). This implies additional terms in the Hamiltonian equations, and has the advantage that it admits natural real space bias models that couple the large-scale dark matter distribution to the galaxy population (see e.g. Wang et al. 2013; Kitaura, Yepes & Prada 2014; Neyrinck et al. 2014; Ata, Kitaura & Müller 2015; Kitaura et al. 2016a; Vakili et al. 2017, and references therein). The only restriction to our approach is to include a treatment for the virialized motions, as we are not able at this moment to include a random term within the analytical posterior sampling expression. Virial motions can, in general, be corrected for with fingers-of-god (FoGs; Tully & Fisher 1978) compression (Tegmark et al. 2004; Tully 2015; Tempel et al. 2016; Sorce & Tempel 2017). Within a Bayesian context, it is possible to include a virialized cluster term following the approach of Kitaura (2013), which includes a likelihood comparison step object by object and not cell by cell as in Jasche & Wandelt (2013) or Wang et al. (2013). In the object-based approach, the likelihood comparison can be done in RSS, including virialized motions as was demonstrated in Heß, Kitaura & Gottlöber (2013). This method requires two steps within an iterative Gibbs-sampler to handle the reconstruction problem, one to transform dark matter tracers from RSS at final cosmic times to real space at initial cosmic times, and a second one to make the link between the discrete distribution to the continuous Gaussian field. It is, in fact, also possible to include an additional Gibbs-sampling step in the Bayesian formalism to self-consistently sample peculiar motions and solve for RSD (Kitaura et al. 2016b; Ata et al. 2017).

In this work, we propose a self-consistent transformation of the *data model* into RSS, allowing to sample directly from the posterior distribution function *in one step* within a Hamiltonian Monte Carlo sampler. Thus, this represents the first self-consistent (coherent) RSS data based primordial density fluctuation reconstruction code based on one iterative sampling step, which can also naturally deal with masks. Our implementation of this approach is called BARCODE. A high-level overview of the method and the problem it solves is presented in Fig. 1. In this code, we transform from Lagrangian coordinates \mathbf{q} to Eulerian coordinates \mathbf{x} using a structure formation model and within the same equation also from \mathbf{x} to the corresponding RSS \mathbf{s} . To implement the RSS transformation into the Hamiltonian Monte Carlo formalism, we need to derive the corresponding analytical derivatives. For our structure formation model, we will restrict our studies to the ZA, as we seek the most efficient, simple formalism, which has been shown to be accurate on scales down to a few megaparsecs when it is fully treated (see e.g. Monaco et al. 2013; White 2015).

1.3 Astrophysical motivation

BARCODE has already been referred to in various works giving details of its implementation (Bos 2016; Bos et al. 2016). One of our main motivations for developing BARCODE is the analysis of galaxy clusters for which the mass has been measured in X-ray (Sarazin 1986; Ikebe et al. 1996; Mulchaey 2000; Nulsen, Powell & Vikhlinin 2010), Sunyaev–Zel'dovich (SZ) cluster surveys (e.g. Birkinshaw 1999; Bleem et al. 2015; Planck Collaboration XXIV 2016; Hilton et al. 2018) and weak lensing studies of clusters (e.g. Kaiser & Squires 1993; Lee et al. 1997; Umetsu 2010; Laureijs et al. 2011; de Jong et al. 2012; Hoekstra et al. 2013; Sartoris et al. 2016; Radovich

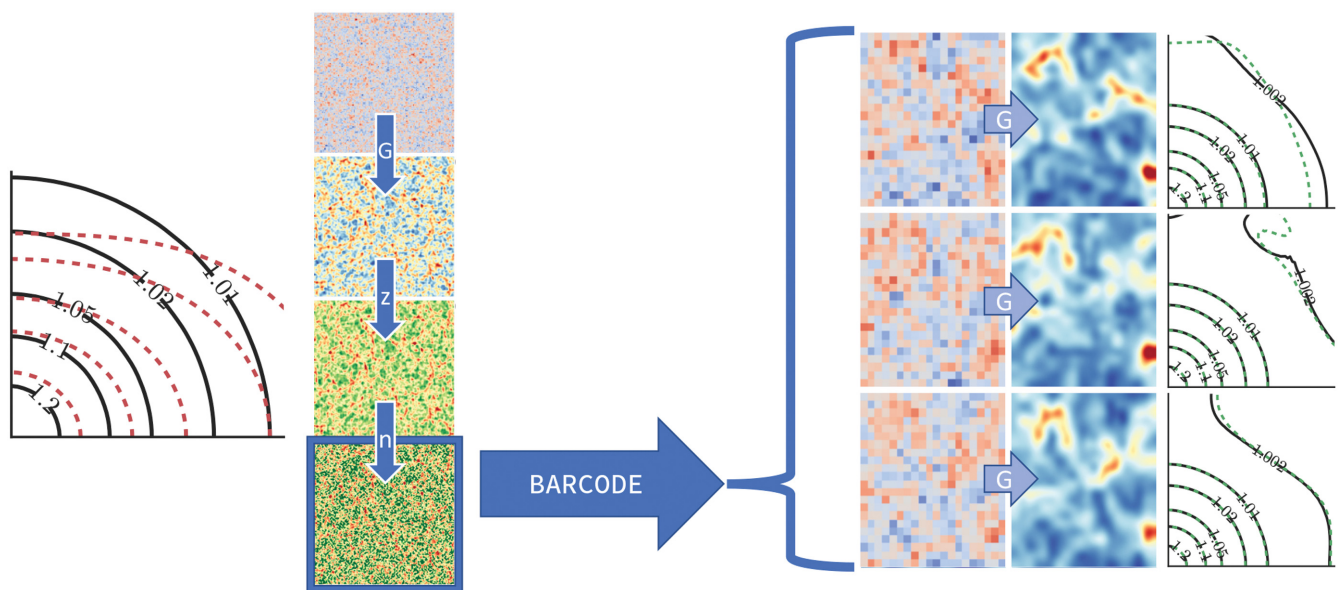


Figure 1. A high-level overview of the BARCODE workflow and pipeline. Left-most panel: illustration of the challenge posed when applying a reconstruction algorithm on redshift space distorted input data. The 2D correlation function of an (statistically) isotropic density field is a nearly perfect circular pattern. This is what we see in the solid black lines that represent the expected (true) field that we want to reconstruct. When comparing the 2D correlation function of the real – non-redshift – space dark matter density fields without correcting for the RSDs, the result is a strong anisotropic deviation (the red dashed lines) from this expected pattern. BARCODE has been specifically designed to deal with this circumstance, and samples primordial density fluctuation field realizations that are constrained to evolve into a present-day mass distribution in accordance with the input set of observations. BARCODE seeks to reconstruct the underlying real-space dark matter density field from a set of observational input data. The procedure is statistically sampling the allowed realizations under these observational constraints. A crucial distinguishing feature of BARCODE is its ability to directly process observations as they are obtained in redshift space without the need for an intermediate redshift distortion correction step. Also implicit in the formalism is that it takes into account the stochastic nature of the relation between the observational input data and the underlying density field. The four left-hand side panels illustrate this sequence of complications. Top panel: the primordial density fluctuations that we sample. Second panel: the corresponding dark matter density field in real space, gravitationally evolved from the primordial field. Third panel: the field in redshift space. Bottom panel: the redshift space field with ‘observational’ noise added. The BARCODE procedure described in this work produces likely reconstructions of the density field, on the basis of the physical and statistical assumptions and prescriptions of the Bayesian model relating the observed mass distribution to the primordial density perturbations. It takes into account the gravitational evolution of the mass distribution, stochastic processes, and observational effects. The right-hand side panels show zoom-ins on to the three realizations of the BARCODE procedure: primordial fields on the left, corresponding gravitationally evolved real space fields on the right. Note that we sample primordial fields, which can in turn be used as initial conditions for evolution studies of the large-scale structure. The images reveal the substantial level of variation between the resulting reconstructions of the underlying dark matter distribution. The differences between the reconstructions reflect the uncertainties emanating from the biased and noisy observational input data, as well as those induced by the approximate nature indigenous to the physical modelling of gravitational evolution and biasing. Despite the intrinsic variations between permissible distributions of the dark matter distribution, the success of the BARCODE procedure in correcting for the redshift distortions in the input observational data is evidenced by the three far right-hand side panels. For three different input fields, the algorithm has produced near-perfect isotropic correlation functions in real space, clear evidence for the removal of anisotropies due to presence of redshift distortions.

et al. 2017; The LSST Dark Energy Science Collaboration 2018). This frees us from having to consider galaxy bias, contrary to many similar studies in the literature (see e.g. Birkin et al. 2019). Rather, in this work we deal with cluster bias, which is very different from the classical galaxy bias. In the case of galaxies, we have tracers and a bias factor (in the linear case) to roughly relate galaxy density and the underlying full density field. In our case, we want to identify the peaks in the density field and ‘move them back in time’ to recover the full underlying surrounding primordial density field; this is, in fact, the central focus of this work. The peaks of the density field are biased tracers according to the background split picture introduced by Kaiser (1984). This implies that having unbiased estimates of the density distribution of a cluster from, e.g. X-ray or lensing measurements, will not lead to unbiased estimates of the whole dark matter field, unless the density field in the intracluster region is properly reconstructed. This means that we are facing two different problems in terms of bias. One is to solve for the bias within the cluster and another is to solve for the bias introduced from having density estimates only in cluster regions. The first problem can

be solved either by preparing the data beforehand (see e.g. Wang et al. 2009) or with a self-consistent deterministic and stochastic bias description within a Bayesian framework (Ata et al. 2015). The second problem can be also solved with a deterministic and stochastic bias description or alternatively with a mask. In the latter approach, the regions where no cluster data are available can be set to have a zero response function, whereas the regions where the cluster has been observed can be set to one in the response function (see e.g. Zaroubi et al. 1995; Kitaura & Enßlin 2008; Jasche & Kitaura 2010; Wang et al. 2013). We have taken this latter approach. Once the full dark matter field is reconstructed, the results are unbiased (see results in Section 4).

More details to the context and motivation of our focus on clusters have been given in Bos (2016). We note that in such a context, the approach proposed here is potentially more appropriate than the ones cited above modelling tracers as point-like objects because the shapes and orientations of clusters play a decisive role in their tidal effects on their environments (Bond et al. 1996; van de Weygaert & Bertschinger 1996). In practice, we will in most cases be forced to

limit our orientation constraints to the two coordinates on the sky, the projected axes of the full ellipsoid. In rare cases, rotation measure tomography (Pizzo et al. 2011) or gravitational lens reconstruction (e.g. Lagattuta et al. 2019) may provide additional information about the radial structure. We leave a more detailed investigation of this to later work.

Broader cosmological applications are also envisioned. Extending this work to make growth rate sampling is trivial, and can be done in a similar fashion to Granett et al. (2015), but including a more complex and accurate algorithm, as presented here. In fact, RSDs can also be leveraged as a proxy for constraining the nature of gravity and cosmological parameters (e.g. Berlind, Narayanan & Weinberg 2001; Zhang et al. 2007; Guzzo et al. 2008; Jain & Zhang 2008; Nesseris & Perivolaropoulos 2008; McDonald & Seljak 2009; Percival & White 2009; Song & Koyama 2009; Song & Percival 2009; White, Song & Percival 2009; Song et al. 2010; Zhao et al. 2010; Song et al. 2011). To this end, many recent studies focus on the measurement of RSDs (Cole, Fisher & Weinberg 1995; Peacock et al. 2001; Percival et al. 2004; da Ângela et al. 2008; Guzzo et al. 2008; Okumura et al. 2008; Blake et al. 2011; Jennings, Baugh & Pascoli 2011; Kwan, Lewis & Linder 2012; Okumura, Seljak & Desjacques 2012; Reid et al. 2012; Samushia, Percival & Raccanelli 2012; Blake et al. 2013; de la Torre et al. 2013; Samushia et al. 2013; Zheng et al. 2013; Bel et al. 2014; Beutler et al. 2014; Okumura et al. 2014; Samushia et al. 2014; Sánchez et al. 2014; Tojeiro et al. 2014).

1.4 Outline

This paper is structured as follows. First, we present the methodology, giving core high-level details of the analytical implementations (see Section 2), while leaving most mathematical derivations to the appendices. In particular, we focus on the novel RSS formalism implemented in BARCODE. In Section 3, we set the stage for our analyses by describing and illustrating our mock observations and the reconstructions that BARCODE yields from them.

The core results of our investigation are described in Section 4. The reconstructions including the RSS model demonstrate the major benefits of our novel approach. For contrast, we show what would happen if one used BARCODE without our RSS extension on observational data in RSS coordinates in Appendix C.

We close with a discussion of the results in Section 5 and conclusions in Section 6.

Additional material is available in the appendices. We first briefly go into more detail than in Section 2 about the most important BARCODE concepts and equations in Appendix A. Then, in Appendix B we present the full derivation of the necessary formulas for implementing our RSS model in the HMC formalism. The astronomically oriented results of Section 4 are followed by a more technical discussion on the performance aspects of BARCODE with this new model in Appendix D. There we evaluate whether the added stochasticity that RSS transformations add cause the MCMC chain to evolve differently than without our RSS model. The 2D correlation function, which is used extensively in the results sections, is defined in Appendix E. An important technical aspect in BARCODE is the use of kernel density estimation, the details of which we discuss in Appendix F. Finally, in Appendix G, we give a complete derivation of the equations necessary for configuring BARCODE to use a second-order Lagrangian perturbation theory (2LPT) structure formation model, optionally including the spherical collapse (SC) terms of the ‘ALPT’ model.

2 METHOD

The Bayesian reconstruction algorithm connecting the evolved dark matter density to the primordial fluctuation field has already been presented in several works (Kitaura et al. 2012; Jasche & Wandelt 2013; Kitaura 2013; Wang et al. 2013; Bos et al. 2016). We summarize the core concepts used in BARCODE – our C++ implementation of this algorithm – in Section 2.1 and list and derive central equations in Appendix A.

The central tenet of this paper is that now we consider the Hamiltonian Monte Carlo calculations in RSS s (equation 2) instead of Eulerian coordinates \mathbf{x} . Densities are then functions of s , e.g. $\rho^{\text{obs}}(s)$. For convenience, in this section we write \tilde{q}_i instead of $\delta(\mathbf{q}_i)$ for the initial density field (the sample in the MCMC chain), where i is a grid location (i.e. a component of the sample vector).

In Section 2.2, we will define what we mean by RSS coordinates. Next, in Section 2.3, we give a brief overview of the different kinds of non-linearities involved in the RSS transformation and our LPT models of structure formation. In particular, we discuss what this means for the representation of non-linear cosmic structure.

2.1 BARCODE

Before we describe our RSS formalism within the Hamiltonian Monte Carlo reconstruction method, we briefly recap the basic idea behind BARCODE, our implementation of this method. BARCODE is a code that aims at reconstructing the full Lagrangian primordial density perturbation field corresponding to a configuration of observed objects. It has been designed specifically for clusters of galaxies as principal constraining objects.

We want this reconstruction to reflect the inherently stochastic nature of both observations and the non-linear structure formation process. This stochasticity can be encoded in a *probability distribution function* (PDF), or posterior. The posterior describes the complete physical process leading to an observed density field; starting from the initial density perturbations, going through gravitational collapse and finally including observational effects.

BARCODE uses an MCMC method called Hamiltonian Monte Carlo (Neal 1993; Taylor, Ashdown & Hobson 2008; Jasche & Kitaura 2010; Neal 2012) to *sample* this posterior, comparing the results to the actual observed density field and in this way finding the initial density distributions that match the data. This yields not just one reconstruction, but an ensemble of possible reconstructed Lagrangian density fields. The variation in these realizations reflects the full array of stochastic effects as included in the model.

To compare Lagrangian space primordial density fields (our ‘sample variable’) to the observations at $z = 0$, BARCODE evolves the sampled Lagrangian fields forward in time using a structure formation model. Then, in the likelihood (the Bayesian posterior being the product of likelihood and prior), the difference of the evolved sampled field to the observed field is quantified. The likelihood effectively ‘scores’ the sample by its similarity to the observations.

In the HMC method, the PDF is represented by a Hamiltonian (Appendix A1), where the PDF’s stochastic variable (the signal we sample) is treated as a position analogue. The HMC algorithm then samples the posterior PDF by repeating the following procedure:

- (i) *Initialize the signal*: define the ‘starting position’ in the parameter space (either a manually supplied initial guess or the position from the previous sample). Our ‘signal’ $\delta(\mathbf{q})$ is defined as

$$\delta(\mathbf{q}) \equiv D_1 \Delta^{(1)}(\mathbf{q}), \quad (1)$$

where D_1 is the first-order growing mode linear growth factor and $\Delta^{(1)}(\mathbf{q})$ is the spatial part of the first-order term of the LPT primordial density field expansion.

(ii) *Stochastic step*: draw a random ‘momentum’ vector (we use a Gaussian field momentum distribution with the Hamiltonian mass as the covariance matrix).

(iii) *Dynamical step*: solve the Hamiltonian equation of motion using a leap-frog solver (Appendix A2) to find the next ‘position’, i.e. the next sample *candidate*.

(iv) *Accept or reject* the candidate based on the acceptance criterion (Appendix A3).

This process can continue indefinitely. Usually, the user supplies some maximum number of desired samples.

In Appendix A, details can be found on our HMC formalism, its terminology, and the algorithmic steps listed above. For the main discussion of this paper, it suffices to mention that the so-called *Hamiltonian likelihood force* plays a central role in the algorithm. Without going into too much detail here, this quantity, which is defined as the derivative of the likelihood to the signal, is used to find the next sample in our HMC Markov chain. This term is mathematically non-trivial, which is why we devote Appendix B to its derivation. It is also computationally quite intensive, which calls for a highly efficient implementation. However, it is of critical importance since it provides HMC with the information from both the data, its stochastic properties, and our assumed structure formation model. HMC uses this information to efficiently ‘orbit’ the high-dimensional space of signals (primordial density fields) and especially to converge quickly on the region in this space of highest posterior probability, i.e. of fields that best match the data.

2.2 Redshift space coordinates

From a practical perspective, we will describe RSS as follows. RSS coordinates \mathbf{s} can be defined as

$$\mathbf{s} \equiv \mathbf{x} + \frac{1}{Ha} ((\mathbf{v} - \mathbf{v}_{\text{obs}}) \cdot \hat{\mathbf{x}}) \hat{\mathbf{x}} \equiv \mathbf{x} + \Psi_s, \quad (2)$$

where \mathbf{s} is a particle’s location in RSS, \mathbf{x} is the Eulerian comoving coordinate, \mathbf{v} is the particle’s (peculiar) velocity and \mathbf{v}_{obs} is the velocity of the observer, and $\hat{\mathbf{x}}$ is the comoving unit vector. a is the cosmological expansion factor, which converts from physical to comoving coordinates, and H is the Hubble parameter at that expansion factor, which converts the velocity terms to distances with units $h^{-1}\text{Mpc}$. We call Ψ_s the *RSS displacement field*, analogously to the displacement field Ψ that transforms coordinates from Lagrangian to Eulerian space in LPT (equation A16).

One important point to stress is that in equation (2) and all equations henceforth, the comoving coordinates are defined such that the observer is always at the origin. The unit vector $\hat{\mathbf{x}}$, and thus the entire RSS, changes completely when the origin changes.

From redshift measurements z , we can derive the radial component of the total velocity \mathbf{u} of galaxies and other objects:

$$\mathbf{u} = \mathbf{v}_H + \mathbf{v}, \quad (3)$$

where \mathbf{v}_H is the Hubble flow velocity and \mathbf{v} is the peculiar velocity. RSS (comoving) coordinates \mathbf{s} can be written in terms of the total velocity as

$$\mathbf{s} \equiv \frac{\mathbf{u}}{Ha} \cdot \hat{\mathbf{x}} \hat{\mathbf{x}}. \quad (4)$$

We can compare such observational RSS positions directly with theoretical models by transforming our models to RSS as given in equation (2).

2.3 RSDs in Lagrangian perturbation theory

For the results presented in this paper, we used first-order LPT – better known as the ZA – as our model of structure formation (see Appendix A4).

LPT is accurate in its position and velocity estimates up to (quasi-)linear order. Note that since our objective is to use them in a model based on data at low redshift, and we are dealing with strong non-linearities, LPT is not a fully accurate model.

To clarify, there are two types of non-linearity involved in RSD theory (Yahil et al. 1991; White, Carlson & Reid 2012). Kaiser (1987) and others – when talking about linear RSDs – are dealing with non-linearities in the mapping from Eulerian to RSS. One marked example of this non-linear mapping is the creation of caustics in the triplen value regions. In the centres of clusters, we also have to take into account that the velocity field itself is non-linear, which, for instance, leads to FoGs.

Since we make no approximations in the RSS mapping, we are dealing with its full non-linear form. The (quasi-)linear velocity field of LPT, on the other hand, will become increasingly inaccurate as we venture into the non-linear density and velocity-field regime, leading to increasingly inaccurate RSS representations.

Large-scale (linear regime) coherent flows are still well modelled in our LPT-based framework. The large-scale velocity field \mathbf{v} in LPT scales linearly with the displacement field Ψ :

$$\mathbf{v} = aH \sum_{i=1}^o f^{(i)} \Psi^{(i)}, \quad (5)$$

where i is summed up to the used perturbation order o ($o = 1$ for the ZA and $o = 2$ for 2LPT) and f is the linear velocity growth factor:

$$f^o \equiv \frac{\partial \ln D^o}{\partial \ln a}, \quad (6)$$

In linear theory, the displacement is directly proportional to the gravitational acceleration, which follows directly from the initial density field. On large scales, structure still resides mostly in the linear regime, so that LPT manages to describe the velocity field fairly accurately. The squashing effect on large scales will therefore be well represented in our model.

Cluster in-fall is modelled poorly in LPT. Virialized and multi-stream motion are not at all included in LPT. This means we do not model FoGs nor the triple value regime within the turnaround radius (see e.g. van Haarlem & van de Weygaert 1993) in this work. Any coincidental triple value regions cannot be trusted in our approach, even though there might be FoG-like features around clusters in LPT as well. Since LPT breaks down there, they cannot be considered as true features. In any case, only about 30 per cent of the FoGs can be attributed to coherent motions, the rest being the result of random motions (Zheng & Song 2016). Even with a prohibitively more expensive structure formation model that could reproduce such motions, a statistical description of FoGs is necessary if no peculiar velocity information is available, as the phase space dynamics in the Vlasov equation cannot be solved. We shortly come back to this in the discussion in Section 5.2.6.

In this work, we use a plane parallel approximation of RSS. Effectively, we put the observer at an infinite distance. This greatly simplifies the equations derived for the Hamiltonian likelihood force in Appendix B.

Table 1. Cosmological parameters used in this work. These are results from the WMAP 7-yr observations (Komatsu et al. 2011), except for the box size.

	Value	Description
Ω_m	0.272	Matter density parameter
Ω_b	0.046	Baryonic matter density parameter
Ω_Λ	0.728	Dark energy (DE) density parameter
w	-1	DE equation of state parameter
n_s	0.961	Primordial power spectrum power-law index
w_a	0	DE linear time dependence parameter
σ_8	0.807	Density field fluctuation rms at $8 h^{-1}$ Mpc
h	0.702	Hubble parameter (units of $100 \text{ Mpc km}^{-1} \text{ s}$)
L	200	Comoving box size in h^{-1} Mpc

3 INPUT DATA

Our novel approach towards integrating RSS in the Hamiltonian solver code needs to be validated. The question of whether the code converges correctly towards the true Lagrangian density and in what way it diverges from it (if differently than in the regular Eulerian space treatment of previous works) will be answered in Section 4 and Appendix D. To this end, we sampled a set of reconstructions based on mock observed density distributions. We describe and illustrate the mocks and reconstructions in the rest of this section.

In our numerical experiments, we use four categories of parameters, detailed in the following subsections:

- (i) cosmological (Section 3.1),
- (ii) statistical (Section 3.2),
- (iii) astronomical (Section 3.3), and
- (iv) numerical (Section 3.4).

3.1 Cosmological parameters

For the cosmological parameters, we use the maximum likelihood results from the WMAP 7-yr observations (Komatsu et al. 2011).¹ The relevant parameters for BARCODE are given in Table 1. The cosmological parameters also determine the power spectrum. We use CAMB (Challinor & Lewis 2011) to generate the power spectrum that is used for the prior computation in BARCODE. We consider cubical volumes of $1.95 h^{-3} \text{ Gpc}^3$. The Zel'dovich structure formation model is applied to transform from Lagrangian to Eulerian coordinates at $z = 0$.

3.2 Statistical parameters

BARCODE has a number of parameters to tune HMC and some choices of statistical models. The results in this paper use the following statistical parameters:

Leap-frog step size ϵ : We use an adaptive step size (Appendix A2.1). It is bounded to be between 0 and 2, but within those limits it is adapted to give an acceptance rate of 0.65 ± 0.05 (calculated over the 50 previous steps).

Number of leap-frog steps per iteration N_ϵ : The number of leap-frog steps is randomly drawn from a uniform distribution of integers between 1 and 256. For Appendix D3, the maximum was increased to 4096 to assess performance as a function of this parameter.

Hamiltonian mass: In HMC, the *Hamiltonian mass* in the kinetic energy of equation (A2) is used as the inverse of the correlation matrix of the Gaussian distribution from which the Hamiltonian

momenta are drawn (see Appendix A for technical details). A smart choice of mass enhances Markov chain performance because it will tune the magnitude of momentum components to their respective *Hamiltonian position* dimensions (Neal 1993). Recall in this that Hamiltonian position equals our N-dimensional signal, i.e. the gridded primordial density field. We use the inverse correlation function type mass (see Taylor et al. 2008; Bos 2016, for more details):

$$\hat{\mathcal{M}} = 1/P(k). \quad (7)$$

This makes sure that the scale of momentum fluctuations matches that of the primordial density field, though only on a statistical level, i.e. without taking into account the actual spatial structure of the sampled field at that point in the Markov chain.

Likelihood: We use a Gaussian likelihood as described in equation (A5). For this work, we set the parameters of that equation σ_i and w_i to 1 for all cells, effectively removing these parameters from the algorithm. When needed, these parameters can be used to quantify per cell i , respectively, the amount of uncertainty in the observed density and a window function to weigh regions of the observed density field more or less strongly. As described in the next subsection, we generate the noise (right-hand panels of Fig. 2), so σ_i is given.

3.3 Astronomical/observational parameters

Mock observed density field: The algorithm is based on the comparison of an observed density field with sampled fields from the posterior PDF. For the runs in this study, we generated the mock observed fields in RSS (instead of Eulerian space as in most previous works (Jasche & Wandelt 2013; Wang et al. 2013; Bos 2016)) as follows. First, we generate a Gaussian random field given the cosmological power spectrum. This Lagrangian space field is mapped to an Eulerian space density field at $z = 0$, using the assumed structure formation model, the ZA. Without fundamental changes, the formalism can be upgraded to higher order approximations such as 2LPT, as shown in Appendix G. At the cost of increased complexity of mainly the Hamiltonian force (Section 2.1), an N-body method could be used as well, as first shown in Wang et al. (2014). Here, we restrict ourselves to the easily interpreted and implemented ZA. The Eulerian space field is, in turn, transformed to RSS on the basis of the ballistic velocities according to the ZA. Following this, to simulate actual data, we add random noise to the RSS density field. The noise is drawn from a Gaussian with zero mean and σ_i standard deviation.

Note that since our likelihood is Gaussian, we must also allow negative densities to be produced by the noise. In Bos (2016), we truncated negative values so as not to generate unphysical negative masses. This turned out to be the cause of the power deficiency in the reconstructions we discussed at length in that work. In this work, we show the potential of the formalism, including its ability to converge on its target, which is why we allow unphysical negative densities in order to satisfy our model choice of a Gaussian likelihood. In a more realistic use case, using real observed data, one would have to choose an error model that does not allow negative values, such as a gamma or lognormal distribution.

The lower right-hand panel of Fig. 2 displays the resulting noisy mock observed field (the panel above that describes the end result of the mock field used in a model where density is defined in regular comoving coordinates instead of RSS).

For our comparison study, two types of runs were set up, in addition to the runs from Bos (2016) (in this work referred to as

¹For more up to date parameters, see Planck Collaboration VI (2018b).

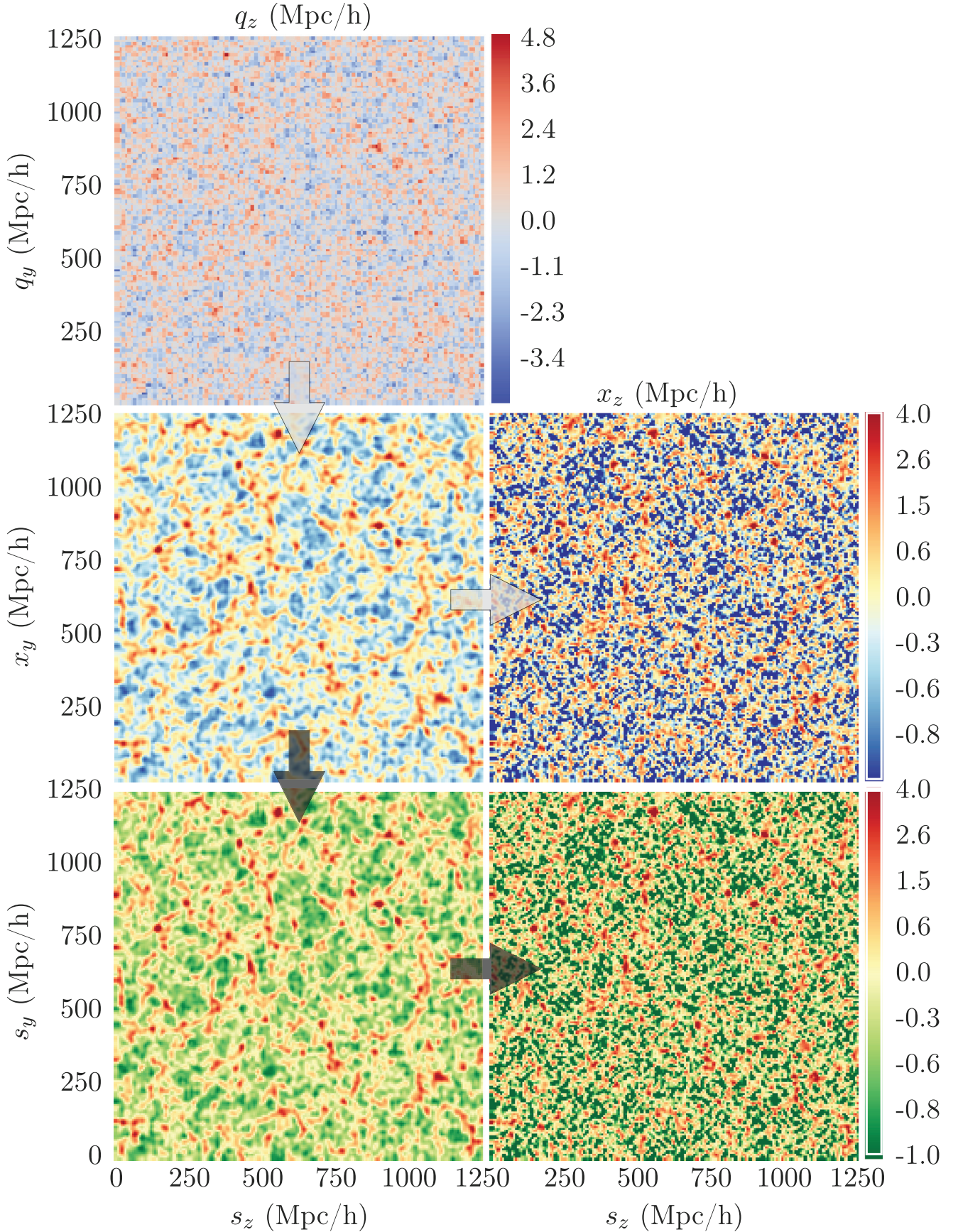


Figure 2. From true Lagrangian field (top left) to true Eulerian field (centre left). From there we can go two ways: either directly to mock observational input field in Eulerian space as we did in Bos (2016) (centre right), or via true redshift space field (bottom left) to mock in redshift space (bottom right).

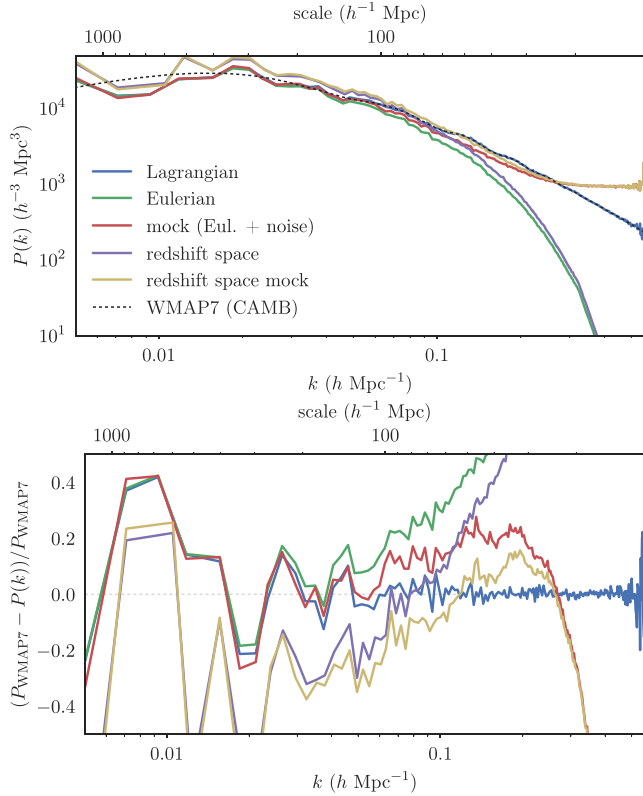


Figure 3. Power spectra of density field in Lagrangian space, Eulerian space, and redshift space and of the mock observational density fields that are used as input (in Eulerian or redshift space). The bottom panel shows the fractional deviation of the spectra from the WMAP7 (CAMB) spectrum that was used to sample the true Lagrangian density field from. The redshift space spectrum peaks above the Eulerian space one, by a factor called the Kaiser factor (Kaiser 1987). The mock observational density fields have a high- k tail that represents the Gaussian noise that is put on top of it.

‘regular’ runs, i.e. runs with data and the model in regular comoving space):

(i) In the first, we used the same algorithm as the runs in Bos (2016), not taking RSS into account; we refer to these as ‘obs_rsd’ runs.

(ii) For the second set of runs, the RSS formalism from Section 2 was used; we call these the ‘RSD’ runs.

In Fig. 3, we show the power spectra of the density fields from Fig. 2. The RSS spectrum clearly reflects the Kaiser effect: a boost of power at low k and a deficiency at higher k .

Window / mask: The window parameter w_i in the likelihood was not used in this work, i.e. it was set to 1 for all cells i .

Starting configuration: To start the MCMC random walk, we need to provide a starting point, or ‘initial guess’, in the N_x^3 -dimensional sampling space. We supply the code with a non-informative guess about the cosmic web structure: a field with value zero in every grid cell. This is also the way to run the code using real observations.

3.4 Numerical parameters

We run with a resolution of $L/N_x = 9.766 h^{-1} \text{ Mpc}$, corresponding to a grid of $N_x = 128$ cells in each direction and a volume side of

$L = 1250 h^{-1} \text{ Mpc}$. The total number of cells is $N = 2097152$. In Bos (2016), we studied computations with $L/N_x = 3.125$. The ZA as a structure formation model correlates well with a full N-body simulation above scales of $\sim 10 h^{-1} \text{ Mpc}$, which is why for this work we chose a higher grid cell size.

We used three seeds to initialize the random number generator in the code. Each of these seeds was used for a separate code run, giving separate, differently evolving chains.

We output the field realizations every 50 steps of the chain. This assures that the saved samples are sufficiently far apart, although it is a bit overcautious, because when the chain performs well, the HMC algorithm should already make sure that subsequent samples are uncorrelated (indeed the likelihood panel in Fig. D2 shows that this is the case). In the statistics of the following sections (means and standard deviations of reconstructions) we use 101 samples, the range from sample 500 to 5500.

We used the SPH density estimator (Appendix F). We set the SPH scale parameter equal to the cell size $L/N_x = 9.766 h^{-1} \text{ Mpc}$.

3.5 Redshift space reconstructions illustration

To illustrate the accuracy of the algorithm, we show and discuss qualitatively the reconstructed density fields. Quantitative results are presented in Section 4. The goal of the algorithm is to recover primordial density fields. However, these are Gaussian random fields with a power spectrum that makes it hard to visually discern large-scale structures. Therefore, instead we compare the evolved density fields to the ‘true’ density field, which illustrates more prominently the accuracy and deviations of the implied large-scale structure.

First, in Fig. 4 we show the difference between the mean and true fields, but now in RSS coordinates s . The mean reconstruction in RSS shows a high degree of similarity to the true field’s configuration, both in the density values and in the shapes of the web structures. The variance around the mean (bottom panel) reflects the propagation of data, model, and systematic errors embodied by the posterior model of the cosmic densities that we sample. As is the case when sampling densities in regular comoving coordinates (see e.g. Bos 2016; figs 3.14 and 3.15), the highest variance can be found in the high-density regions, reflecting the higher number statistics of the particles that cluster to form these regions. These same statistics lead the voids in reconstructions to be dominated by random fluctuations from the prior. These fluctuations average out in the mean field. The underdense voids, hence, contain less structure in the reconstructions’ mean than in the true field.

We can further compare the RSS results to the regular Eulerian space density shown in the top two panels of Fig. 7. The coherent inflow part of the RSDs causes expected features along the line of sight in RSS: squashing and enhancement of density contrast for overdense large-scale structures and a stretching of underdense voids.² These effects are quantified in the power spectrum, which in RSS shows the Kaiser effect (Fig. 3; see also Fig. C3 of Appendix C, which illustrates the imprint of the Kaiser effect on reconstructions that were not corrected for RSDs), and the 2D correlation function, further discussed in Section 4.3.

²These features may be hard to appreciate from static figures. We provide an animated GIF at <http://egpbos.nl/rsd-eul-rsd-rss/> for more convenient comparison of the top right-hand side panels of Figs 4 and 7.

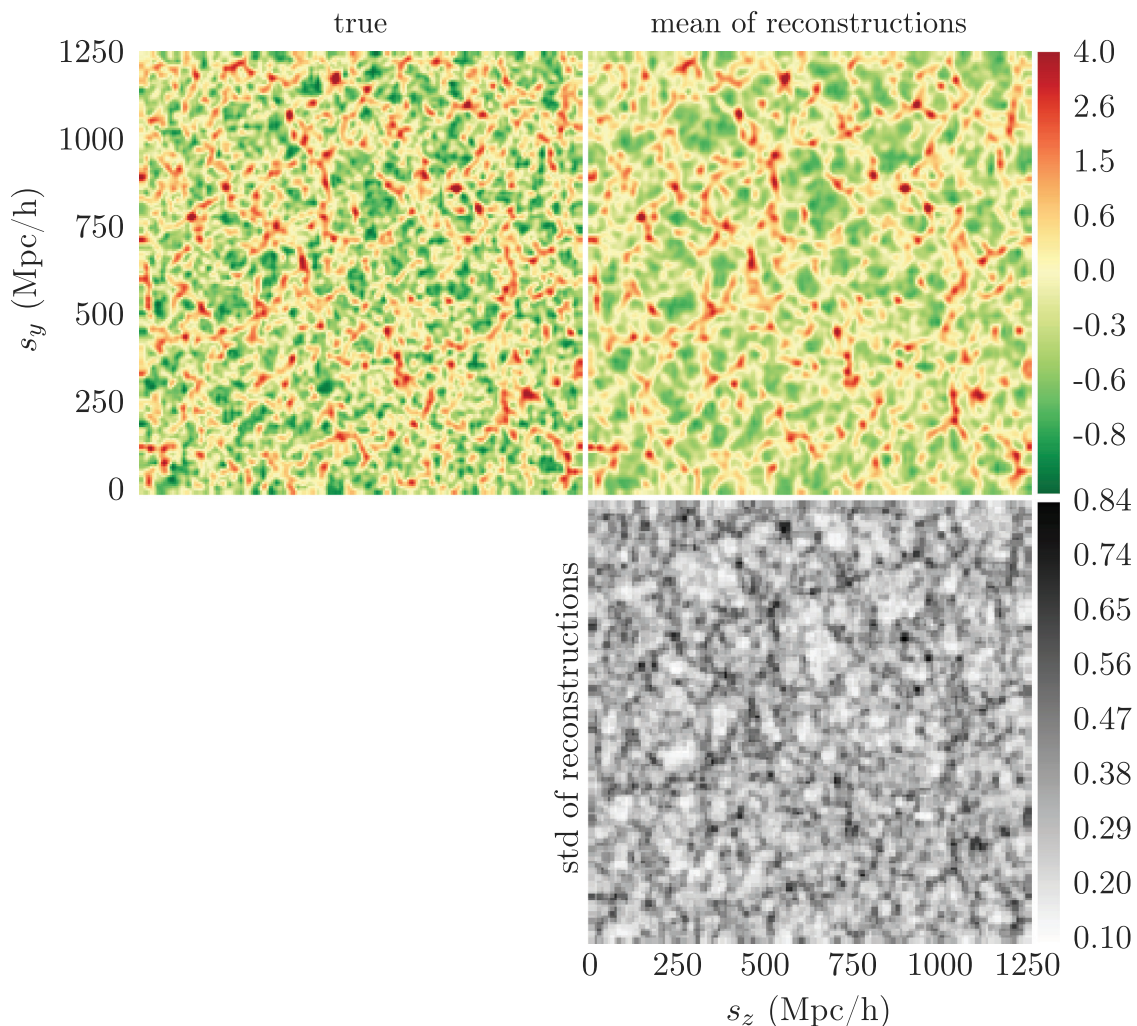


Figure 4. Redshift space density fields. *Top left:* the true field. *Top right:* the mean of the sampled fields. *Bottom:* the standard deviation of the samples.

4 RESULTS

In this section, we will show the benefits of using our self-consistent RSS Zel’dovich model over a regular Zel’dovich model that does not self-consistently treat observations as being in RSS, the latter of which is used in previous models (e.g. Jasche & Kitaura 2010; Jasche & Wandelt 2013; Wang et al. 2013). When properly dealt with, the anisotropies caused by RSDs can be modelled in the data and hence eliminated from the reconstruction. This enables reconstruction of the mean density field based on RSS data. By directly incorporating RSDs in the formalism, ad hoc corrections for RSDs are no longer necessary.

From an algorithmic point of view, the most pressing question is whether the code does converge correctly towards the true Lagrangian density. Also, we want to characterize in what way the samples diverge from it. This latter question is addressed in Section 4.2.

One might wonder what would happen if RSS data are used without an RSS model, i.e. ignoring the effect of RSDs in the observations by not modelling them in the reconstruction algorithm, or more specifically: the posterior. In Appendix C, we illustrate the non-negligible impact of the RSS effects in such a naive approach. This clearly motivates the need for the model used in our approach.

4.1 Large-scale structure reconstruction

We first inspect the match of the Lagrangian density obtained by the chain samples, after burn-in, to the true underlying density. Comparing the true field to the ensemble average in the top panels of Fig. 5, we find a good match in terms of the large-scale structures. Many high peaks and low troughs have their counterpart in the samples, even though their exact shapes and heights do consistently differ. In some cases, peaks or troughs are combined, in some they are split, and in some cases random peaks or troughs are added that should not exist at all. These differences and similarities can be appreciated even better in the zoom-in of Fig. 6.

We should note that these kinds of fluctuations are to be expected. In the individual samples, such fluctuations will be even more pronounced (as we discuss in Section 4.2). In the mean sample, the algorithm is expected to fluctuate around the true density field to a great extent. However, it is highly unlikely that the mean field will ever exactly match the true field. This is prevented by the uncertainties that are inherent to all the processes involved in the posterior.

The algorithm not only manages to reconstruct the structures, also quantitatively it works well: the difference between true and reconstructed density is typically on the same order of magnitude as the values themselves. It leads to a good reconstruction of

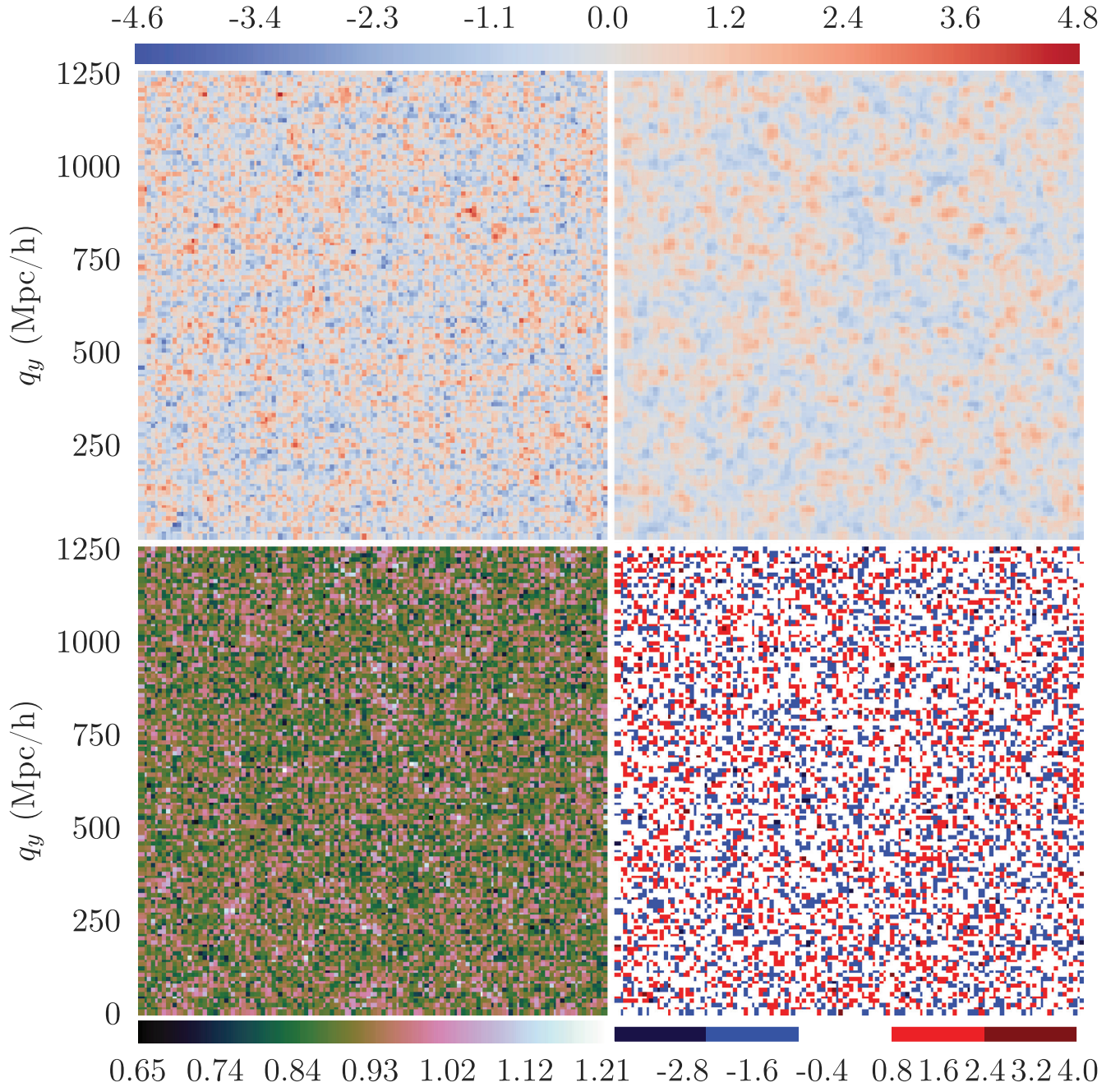


Figure 5. Mean sample *Lagrangian* densities (top right) compared to true (top left), sample std (bottom left) and true minus mean (bottom right).

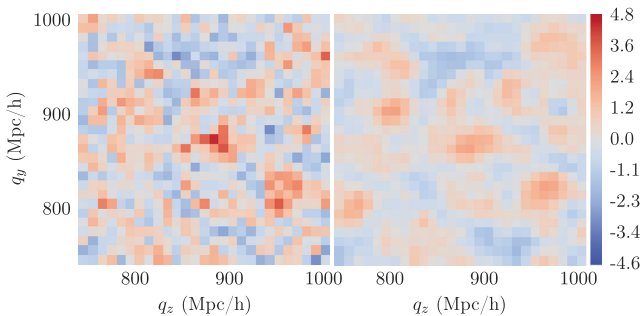


Figure 6. Zoom-in of Fig. 5.

the density distribution function, but still allows for significant variation. The standard deviation of sampled densities is typically around 1 (bottom left-hand panel of Fig. 5). Using the RSS model gives a very comparable amount of variation to when one does not need it, as can be seen by comparing to Fig. C2. In principle, given the added uncertainty in spatial mass distribution configurations along the line of sight introduced by RSDs, one might expect that the variation around the mean would be smaller when using the RSS model. Stated differently, the degeneracy of the z -axis could lead to the ‘leaking’ of the MCMC chain of some of its movement in the statistical coordinate space \tilde{q} to velocity space. Here, we see no such effect, which may indicate that our chosen HMC parameters, in particular N_ϵ , cause

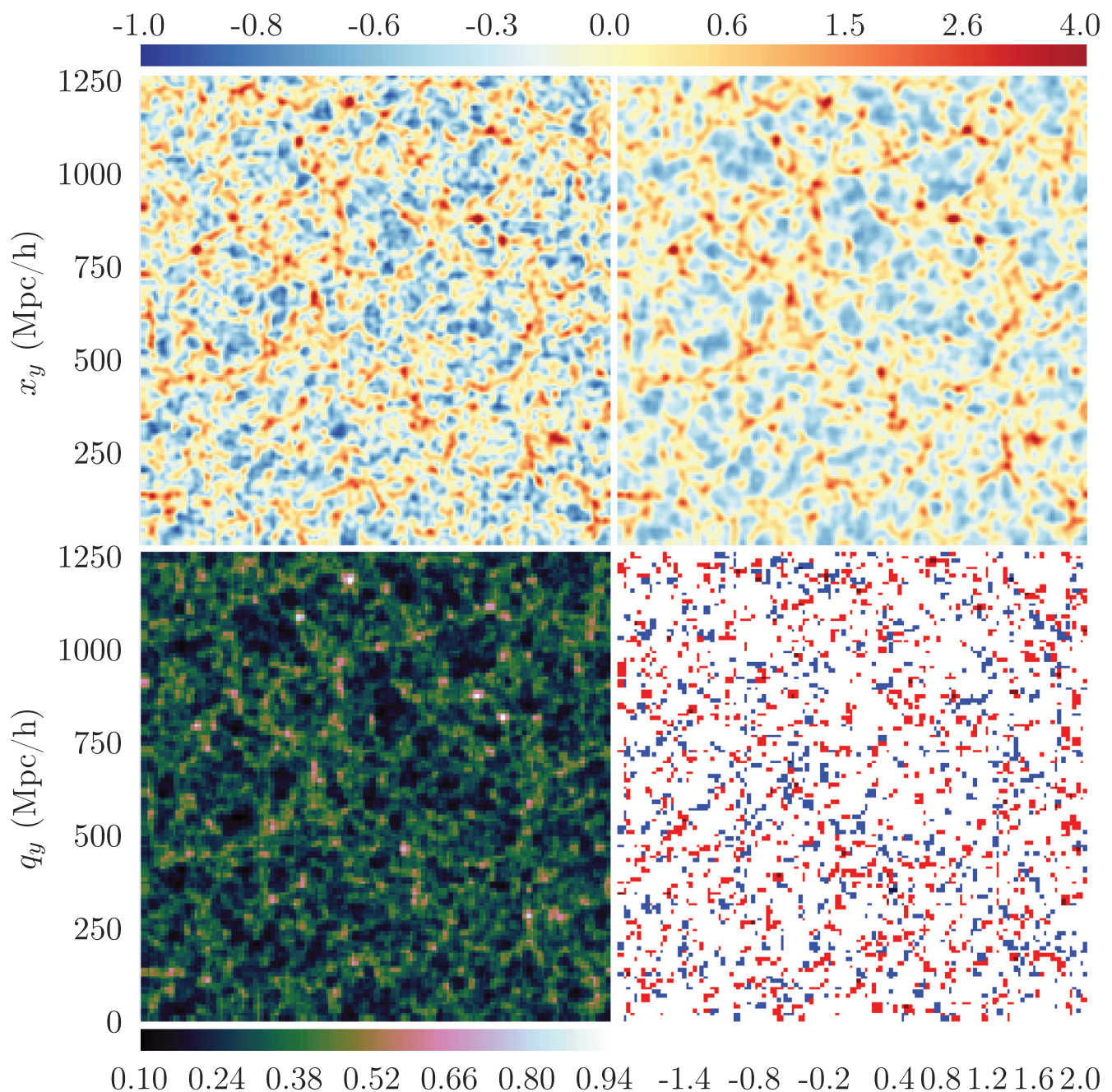


Figure 7. Mean sample *Eulerian* densities (top right) compared to true (top left), together with std (bottom left) and the difference of the true and mean fields (bottom right); with RSD model.

a sufficiently uncorrelated sampling process for both cases. We mention this here explicitly because in our previous work on higher resolution runs (Bos 2016) we did see a difference in sample variability between runs with and without RSS model. The sample variation is further discussed in Section 4.2 and in Appendix D.

Following gravitational evolution, the corresponding Eulerian density fields reveal an evolved web. Our analysis therefore also allows a similarity inspection of the prominent web-like mass distribution features in the cosmic web. On the basis of Fig. 7, it becomes clear that this model performs very well in reconstructing the large-scale features that we are interested in.

As expected, there is still some difference, as seen in the bottom right-hand panel. With the RSDs included in the reconstructions, the outstanding z -directionality or bi-polarity that can be noted in the uncorrected process (see Fig. C2 of Appendix C) have fully disappeared. Moreover, the variation seems evenly spread out over the entire field. If the variation were caused by RSDs, one would expect a concentration of differences around clusters and ‘great walls’. As expected, compared to the Lagrangian case, the standard deviation of the samples is somewhat lower, but of the same order of magnitude. We discuss this further in Section 4.2.

Looking further at the sample power spectra in Fig. 8, we find an excellent match to the expected true spectrum. The difference

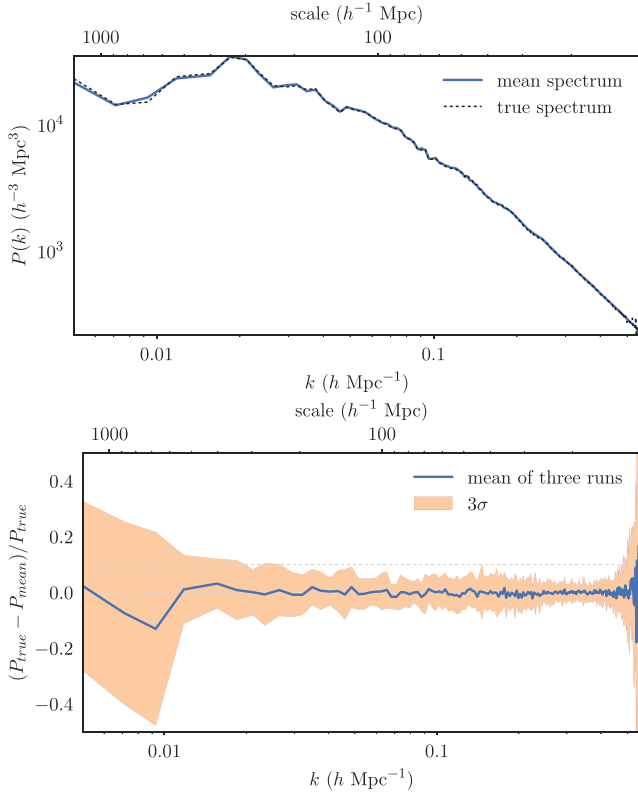


Figure 8. True Lagrangian power spectrum compared to the mean and variance of the sampled density fields’ power spectra *with RSD model*. Note that the difference plot (bottom panel) uses samples from three different runs with different true fields, while the power spectrum plot itself (top panel) is only from one true field.

of the mean to the true spectrum in the bottom panel shows that the Kaiser effect has been completely accounted for. The deficiency in the reconstructed power, peaking at $k = 0.2 h \text{ Mpc}^{-1}$, which we discussed in Bos (2016) is no longer present here, due to allowing the mock observed density distribution to have values $\rho < 0$, so as to properly match the Gaussian likelihood (see Section 3.3).

4.2 Density variation around mean

As demonstrated above, the variation of samples around the mean seems similar to that obtained with a regular Eulerian space model (Appendix C and Bos 2016). One could have expected the algorithm to have somewhat more trouble evolving the MCMC chain in the RSS case. However, as argued above, it seems that in our situation, the leap-frog step size was sufficiently large for the chain to orbit similar volumes of posterior parameter space in both cases.

In Fig. 9, we take another close-up look at the evolution of the chain. The same region of the reconstructed Eulerian density field is shown for six different iterations of the chain in Fig. 9(a). These can be compared to the true density field in the bottom panel. We see here the stochastic sampler at work. The main large-scale features can be identified in all samples, like the very massive peak in the bottom right-hand corner and the filamentary structure in the top left. The massive peak is rather stable, with some additional protrusions sometimes being formed at random. The filaments, however, undergo some major variations in shape, orientation, and connectedness.

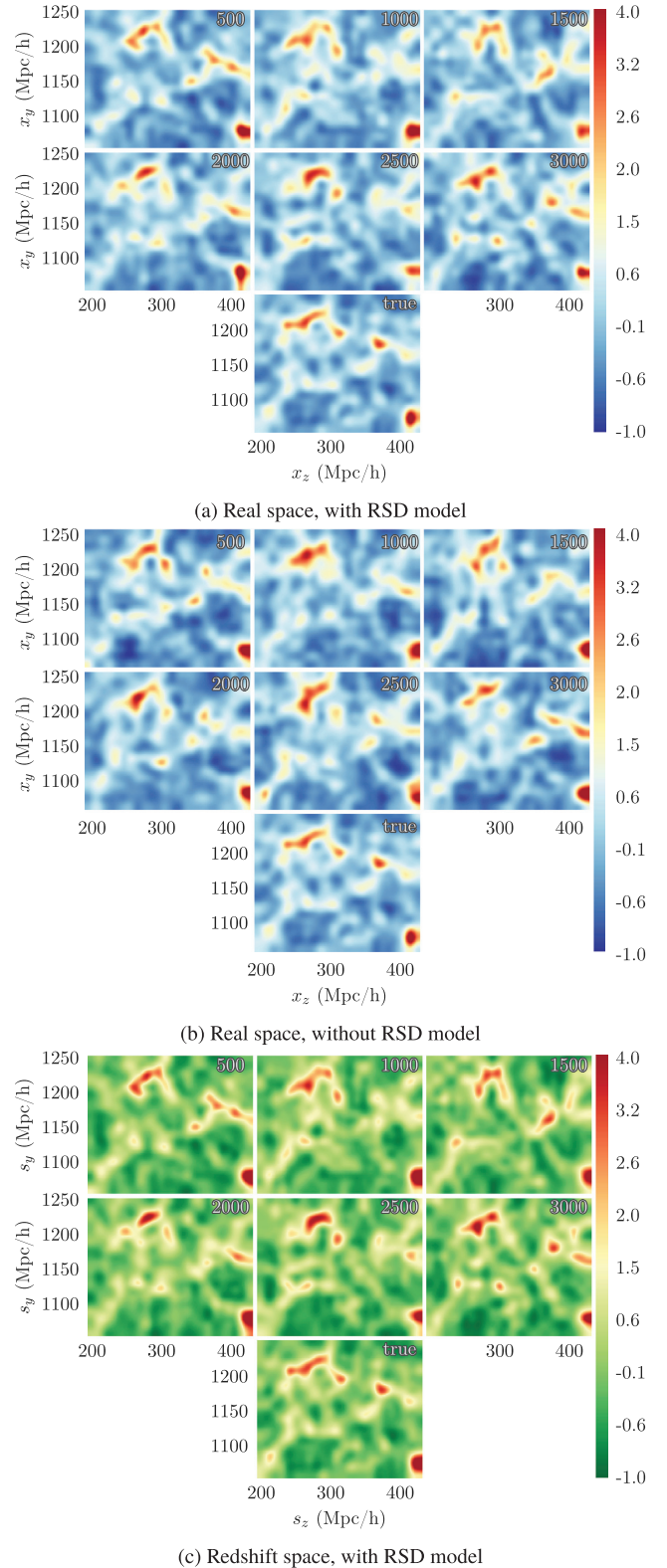


Figure 9. Zoom-in on Eulerian real and redshift space density fields of iterations 500–3000 (top six panels) and the true field (bottom panel): with and without RSD model.

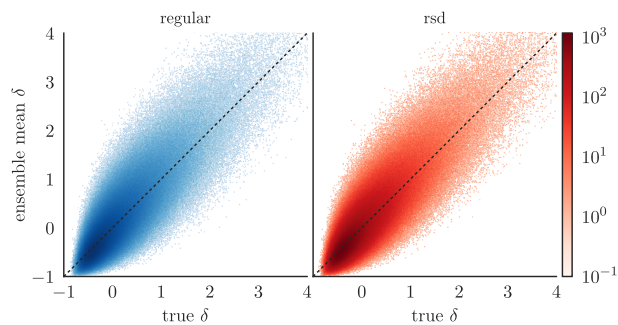


Figure 10. Histogram of the density of the true Eulerian density field (the horizontal axis) versus the corresponding density in the ensemble average of the chain (the vertical axis), compared in each grid cell. In red, the redshift space model and the regular model in blue.

We compare this behaviour to that of a regular space run in Fig. 9(b), see Appendix C for details on this run. While there are clear differences with the RSS-corrected runs of Fig. 9(a), the results do not seem qualitatively different by eye in terms of amount of structural variation, which is in line with what we discussed above. At the same time, we see deeper reds in clusters and blues in voids in Fig. 9(b) than in the corrected runs. This is because the uncorrected runs reproduce features that look like RSD – amongst which: enhanced density contrast – but then in real space. The enhanced contrast can also be seen in the RSS zoom-ins of Fig. 9(c), which corroborates this fact. The uncorrected runs are discussed in more detail in Appendix C.

4.2.1 Density–density full field comparison

Fig. 10 shows, for each grid cell, the density of the true Eulerian density field versus the corresponding density in the ensemble average of the chain. The right-hand panel shows the results for the RSS model, the left-hand panel for the Eulerian space model. These plots corroborate some of our previous findings. The variation (width of the distribution around the $x = y$ line) is more or less the same, corresponding to the similar variation in the samples we saw before. Generally, the correspondence is good for both models.

4.3 2D correlation function isotropy

Because of the anisotropy of RSS, RSD effects show up more clearly in directional statistics such as the 2D correlation function $\xi(\sigma, \pi)$ (see Appendix E for more detail). In $\xi(\sigma, \pi)$, the three spatial directions are collapsed into two. One uses the directions along and perpendicular to the line of sight. Anisotropies in the data will show as a deviation from a perfectly circular $\xi(\sigma, \pi)$. RSDs will thus have a marked signature (see e.g. Peacock et al. 2001; Hawkins et al. 2003; Guzzo et al. 2008).

The 2D correlation function $\xi(r_\perp, r_\parallel)$ of the reconstructions matches the true one when using the RSS model. This indicates that the large-scale RSS anisotropies have been eliminated.³ In Fig. 11, we compare the correlation functions of the ensemble mean of chain samples and the true field. The difference between them, $\xi(r_\perp, r_\parallel)_{\text{true}} - \xi(r_\perp, r_\parallel)_{\text{mean}}$, is maximally ~ 0.001 , with the mean difference being ~ 0.000013 .

³We use the notation $\xi(r_\perp, r_\parallel)$ instead of $\xi(\sigma, \pi)$, replacing $\sigma = r_\perp$ and $\pi = r_\parallel$.

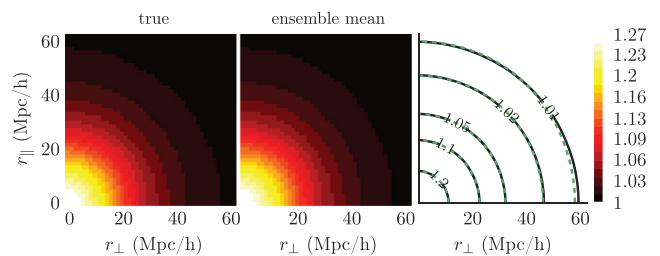


Figure 11. Eulerian space 2D correlation functions: true (left) versus ensemble mean (centre). The right-hand panel shows the two functions in contours: the solid black for the true function and the dashed green for the ensemble mean.

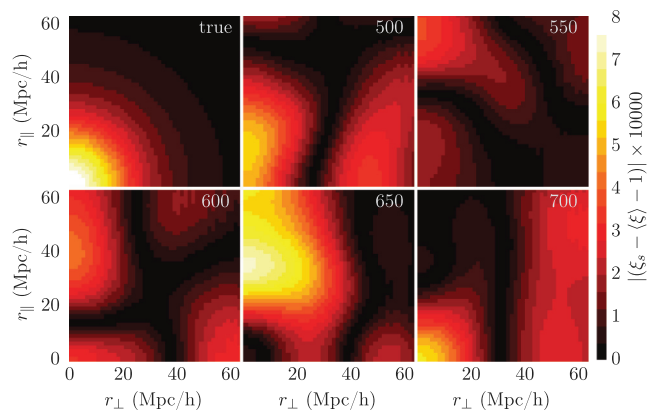


Figure 12. Five sample Eulerian 2D correlation functions minus the mean of all reconstructions, compared to the true field in the top left-hand panel.

The nature of the MCMC sampling process itself introduces additional anisotropies in the samples. No single realization will perfectly match the 2D correlation function of the true field. This sampling effect is illustrated in Fig. 12, where we show the difference of the correlation functions of five individual realizations with that of the mean of all realizations. The differences are very small, on the order of $\delta\xi/\xi \sim 10^{-3}$. To account for this case-by-case variation, we only look at sample ensembles instead of just at single samples.

Note that for reasons of cosmic variance, some amount of anisotropy is always expected. A finite-sized box will never be perfectly isotropic. This fact is illustrated in Fig. 13. The first three rows are each based on a different true Lagrangian density field, generated with a different random seed. It is immediately apparent from their deviation from perfect circularity on larger scales that the true fields (left-hand column) contain a considerable amount of anisotropy. To account for this effect, one should always compare to the true field and take its anisotropies into consideration. Indeed, it can be seen that the true and reconstructed anisotropies match well. Alternatively, one could run a large number of chains with different random input fields and average out their statistics. For the entire Universe, we expect a perfectly isotropic condition, reflecting the cosmological principle. Even averaging over only three different chains does not suffice, as may be appreciated from the bottom row of Fig. 13.

5 DISCUSSION

We developed and tested a self-consistent method for the reconstruction of an ensemble of likely primordial density fields based on mock

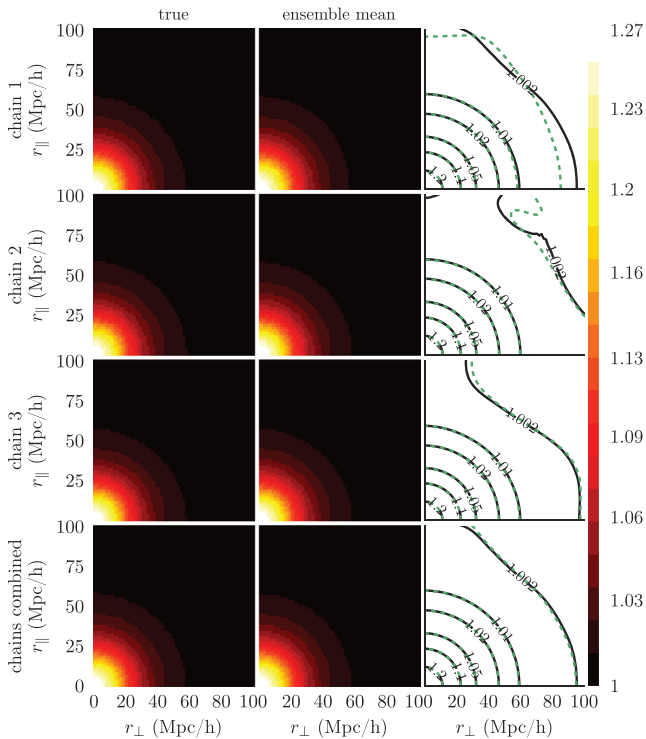


Figure 13. Sample Eulerian 2D correlation functions compared to true: the first three rows are based on three different true Lagrangian density fields, while the last row contains the average of the three runs. The left-hand column shows the true fields, the middle row the mean of the ensemble, and the right-hand row compares the two. The ensemble mean is represented in the dashed green and the true field in the solid black lines. This sample images clearly illustrate the role of cosmic variance in the anisotropy of the field.

observed data in RSS. We showed that this significantly improves the quality of the reconstructions. They are more isotropic and RSS artefacts are removed. This forms a contrast to a naive application of an Eulerian space algorithm to RSS data, which would retain these features.

The novelty of our method relates to the explicit treatment of RSS. We extend existing Hamiltonian MC methods (Jasche & Wandelt 2013; Wang et al. 2013) to a slightly more realistic data model. The rationale is that pre-processing data should be kept to a minimum. Properly modelling the data is always preferable. This way, all the possible interacting complexities in the data will be properly handled, whereas pre-processing data might lead to additional noise and/or errors. For instance, to artificially remove FoGs one must make a lot of simplifying assumptions, such as sphericity of the clusters (see e.g. Tegmark et al. 2004; Tempel, Tago & Liivamägi 2012). Simple, automated FoG removers will inevitably generate false positives and also miss out on true positives. In our integrated RSS approach, no assumptions are needed at all. This is not only considerably easier, but also seems like the only way that one can ever consistently model all the complexities in the data.

Note that redshift distortions have been treated in a different way by Wang et al. (2013). Their treatment of redshift distortions (Wang et al. 2009) is based on a linear approximation of the velocity field, which they use to correct for the positions of dark matter haloes. This is a pre-processing step that is done only once and in a unique way, before running their Hamiltonian sampler. In doing so, the

intrinsic uncertainty from using redshift as a proxy for distance is circumvented rather than accounted for.

In the rest of this section, we will discuss some of the (astro)physical and astronomical applications of our formalism. The method presented in this work is still not fully complete with respect to modelling all complexities of the cosmic structure formation process, which we will also touch upon. In addition, a range of technical aspects and open questions will be addressed and we will discuss some further points pertaining to the novel redshift treatment.

5.1 Regions/environment of applicability

Using the BARCODE framework, we aim to use cluster data as input and statistically study the cosmic web that forms in between. All the stochastic effects involved in the evolution of large-scale structure and in the observations thereof can be self-consistently included. The analysis of the resulting ensemble of reconstructions should then be straightforward. This matter was explored in chapter 5 of Bos (2016).

One point of attention when using this code is the fact that it is biased, and is better tuned towards constraining high-density regions than low-density regions. Because many Lagrangian \mathbf{q} locations will end up in clusters on the real-space \mathbf{x} or \mathbf{s} grid, the real-space constraints in BARCODE are biased towards more high-density regions. This means that even though you might have put constraints on $\rho^{\text{obs}}(\mathbf{s})$ in void-like regions, they will be poorly probed. In a future paper, we will explore this effect in more detail by trying to constrain voids.

Given our aim of using clusters as constraints, this circumstance is actually quite fortunate. One might, however, want to modify the algorithm to better suit void-region constraints. One possibility is to sample directly in Eulerian (volume) space \mathbf{x} , as opposed to the Lagrangian (mass) space. This poses problems with the non-uniqueness of the translation from Eulerian to Lagrangian coordinates. Another option might be to use $\sigma(\mathbf{x})$ to compensate for the voids. One could lower the variance in those regions, thus making the constraints tighter and more likely to be held in a random sample. However, it still leaves the possibility of intermediate fluctuations, given that the amount of particles ending up in voids will be low. The effects of this will also be tested in a future study. This approach is consistent with that of Wang et al. (2013), who define the variance as

$$\sigma(\mathbf{x}) = \mu \delta^{\text{obs}}(\mathbf{x}), \quad (8)$$

with μ a constant, i.e. simply a linear function of the density.

5.2 Future steps towards real data

A number of improvements may be identified that would enable our formalism to improve the accuracy, reliability, and level of realism of our reconstructions. For future work, especially for applying the formalism on real observational data, we here list the most significant ideas towards improvement.

5.2.1 Rank ordered density value correction

When one wants to compare true density fields to fields derived with LPT, one of the first aspects encountered is the difference of one-point density distributions (PDFs). Perturbation theory approaches do not prevent particles from pursuing their initial track. Beyond

shell crossing, it leads to the overshooting of the position they would take if the mutual gravitational interaction would have been taken into account in the calculations. True gravitational clustering will give much higher peak density values than LPT, where this artificial shell crossing artefact creates ‘fuzzy’ clusters. A simple first-order correction to this would be to apply a rank ordered substitution of density values. A transfer function from an ‘LPT Eulerian space’ density to, for instance, a ‘ N -body Eulerian space’ density can be constructed. Leclercq et al. (2013) did this by comparing the rank order of density values in LPT and N -body densities at $z = 0$ (given the same initial conditions) and reassigning densities in one or the other in such a way that the density value PDFs match after the procedure. This way, at least one is comparing apples to apples for as far as first-order statistics are concerned.

5.2.2 Accuracy of gravitational modelling

We have shown that in the perfect situation of a model that exactly describes our (mock) reality, our algorithm almost perfectly reconstructs the desired mean field and its statistics. Although the used models describe cosmic structure well on large scales, they are far from perfect for describing non-linear regime structures. There are a few models that would give a more complete picture of gravitational evolution.

The current LPT (Zel’dovich) based framework is easily adapted to using 2LPT or ALPT (Kitauro & Heß 2013). 2LPT has the advantage of matching even better on the largest scales. At cluster scales, 2LPT is more affected by artificial shell crossing than Zel’dovich, leading to ‘puffy’ structures. The latter can be fixed by combining 2LPT with an SC model on small scales. This is what ALPT accomplishes. Both models are fully analytical, so that they can be implemented in the same manner as the ZA described in this work. In Appendix G, we work out the equations necessary to extend BARCODE with 2LPT and ALPT structure formation models. A similar option would be to apply the Multiscale Spherical Collapse Evolution model (Neyrinck 2016). This analytical model was shown to perform slightly better than ALPT, especially when combined with 2LPT on large scales.

It would be even better if we could use an N -body simulation as our structure formation model. Wang et al. (2014) indeed for the first time successfully invoked a particle mesh N -body algorithm in this context. The particle mesh equations are analytical, so every single particle mesh evolution step can be derived to the signal. By writing the derivative in the form of a matrix operator and combining subsequent particle mesh time-steps by means of matrix multiplications, the full likelihood force can be analytically derived. By means of adjoint differentiation, the large matrices can be efficiently multiplied and the computational cost of this method stays within reasonable limits. The resulting model accuracy using only 10 particle mesh steps is remarkably high. When one needs high accuracy on small scales, this seems like the way forward. More recently, this approach was also adopted by Jasche & Lavaux (2018).

Possibly, the method by Wang et al. (2014) could be augmented using an N -body solver that also has *baryonic particles*. Whether this is analytically tractable within an HMC framework remains to be investigated. Another interesting extension might be to employ the COLA method (Tassev, Zaldarriaga & Eisenstein 2013; Tassev et al. 2015) as an alternative gravitational solver. COLA combines the LPT and particle mesh methods, trading accuracy at small scales for computational efficiency. It yields accurate halo statistics down

to masses of $10^{11} M_{\odot}$, which would be more than sufficient for cluster studies. In fact, the COLA method has already found uses in the context of Bayesian reconstruction (Leclercq 2015; Leclercq, Jasche & Wandelt 2015b), but in these cases COLA was applied after the Bayesian reconstruction step, not self-consistently within the model like in the work of Wang et al. (2014).

5.2.3 Galaxy biasing

As, in practice, observations concern galaxies, stars, and gas, instead of dark matter, it is of key importance to address the issue of in how far the galaxy distribution reflects the underlying dark matter distribution (see Desjacques, Jeong & Schmidt 2018, for a recent review). For the application to clusters observed in X-ray, SZ, or via weak lensing, this is not directly necessary, since for clusters the biasing problem is far less. Schaller et al. (2015) showed that the ratio of halo mass in N -body simulations with and without baryons is $\simeq 1$ from mass $\simeq 10^{13.5} M_{\odot}$ upwards. It drops off towards $\simeq 0.8$ for galaxies. Similarly, the halo abundance in cluster-sized haloes was shown to be similar in simulations with and without baryons. However, we might want to extend the algorithm towards a formalism capable of processing the galaxy distribution (e.g. Wang et al. 2013; Leclercq, Jasche & Wandelt 2015a).

A natural way to solve this would be to incorporate gas particles in the structure formation model. However, this would still be a mere approximation, due to the complexities of baryonic astrophysical processes. A more statistical approach would be to explicitly incorporate biasing prescriptions (see e.g. Ata et al. 2015; Kitauro et al. 2015, and references therein). We have implemented such models in BARCODE and will explore their use further in an upcoming work.

5.2.4 Masking and selection functions

One of the more observationally oriented issues is that of masking and selection functions. In our general Gaussian likelihood (equation A5), we included a weight factor w_i . This can be set to zero if nothing is known about cell i in the data and to one if there is. Another option is to set it to zero if the area is observed, but was found to be empty (at least up to the depth of the observation). Either of these constitute a very basic form of masking.

Given the nature of observations, one could think of more advanced selection mask treatments. For instance, we should take into account the depth of a null-observation and the fact that this implies a *lower limit* to the density, not an absolute absence of mass. One could implement this by setting $w_i = 0$ if $\rho^{\text{obs}} < \rho^{\text{cut}}$ for some cut-off density (per cell). A theoretically more sound model may be to use a Heaviside step function convolved with a Gaussian kernel (i.e. an error function, plus one, divided by two) with a step at the cut-off density ρ^{cut} . This reflects the fact that there is some uncertainty about whether the density is above the cut-off value that we derived from the depth of our observation. It further tells us that we are quite sure it is not much higher and that it is most likely lower than the cut-off value.

Such a selection function will give an additional $\delta(q)$ dependent term in the Hamiltonian force. The derivative is simply a 1D Gaussian distribution. This implies that it should be straightforward to derive and should give only minimal extra overhead in the algorithm.

These masking treatments will be explored further in an upcoming work.

5.2.5 Statistical aspects of the implementation

Another parameter in our Gaussian model we hardly gave any attention is the σ_i cell-wise dispersion of the data. It is highly likely that a more suitable value can be found than the one we used (1 for all cells). In general, the statistical model in the form of prior and likelihood that we use in our formalism might be improved upon. It seems like the prior is well defined and encapsulates what it must in a proper way. The likelihood is a different story.

Other authors have mainly employed Poissonian data models (Kitaura & Enßlin 2008; Jasche & Wandelt 2013; Wang et al. 2013) instead of the Gaussian likelihood we used. The way to find out what model should be used is to meticulously map out the errors and systematics involved in the observations. One particular example of this is that of defining a likelihood model for an observational data set of X-ray clusters. In this case, we have to take into account the following aspects:

- (i) X-ray photon counts, which have Poissonian errors;
- (ii) X-ray photon energies, which are also Poissonian due to CCD photon counts;
- (iii) the propagation of photon errors into a χ -squared fit for the temperature and density profile parameters, e.g. using a β -model (King 1972; Sarazin 1986; Mulchaey 2000) if the confidence ranges of such parameters are estimated with bootstrapping, could lead to complicated likelihood shapes for the density profile parameters;
- (iv) many systematics arising from the use of certain fitting models for density estimation, such as cool cluster cores that are not included in simple β -models (e.g. Schindler 1996); and
- (v) the combination of the density/mass profile errors with errors in the distance estimators, e.g. from redshifts of the brightest cluster galaxies or some weighted average of all clusters' galaxies.

Our Gaussian likelihood was chosen as a general error model that is likely to be close to (but not equal to) many true data likelihood functions. This approach is, in fact, not uncommon in the literature for X-ray cluster density error estimation. In Eckert et al. (2011), an error in the density profile of a cluster is estimated using a Monte Carlo approach, by simulating 10^8 realizations (assuming Poisson statistics) and taking the variation in values as errors. In Samsing, Skielboe & Hansen (2012), an MCMC sampling of the parameter space is done to characterize the probability function around the best-fitting model. Subsequently, the width of the projected PDFs for every parameter were characterized by the root mean square, meaning that essentially they are approximating the error on the parameters as Gaussian, i.e. they leave out higher order PDF characteristics. For the NORAS survey, Böhringer et al. (2000) assume statistically independent, Gaussian-distributed errors, even though they state this is not 100 per cent correct. Finally, many authors use the χ^2 -statistic for determining the goodness of fit in their analyses. This statistic is also based on a Gaussian distribution of noise.

The HMC algorithm is quite sensitive to deviations of the data model from the actual data errors. Using a wrong model can lead to poor reconstructions, as we found in Bos (2016). In that work, we found that the power spectrum was not properly reconstructed, deviating from expected values by as much as 10 per cent in some ranges. For this work, to show the potential of the RSS correction model, we accommodated for the Gaussian likelihood by allowing negative densities in the mock density fields produced by adding Gaussian noise to the Eulerian RSS density field derived from the 'true' Lagrangian initial density field. In other use-cases, when reconstructions deviate significantly

from expectations, one could use this as an indicator of model deficiencies.

5.2.6 RSDs and LPT

The use of LPT imposes two major limitations on an RSD treatment. Because of the inaccuracy of the non-linearities in the velocity field, FoGs and triple value regions are not accurately represented. However, just as clustering in density was augmented in the ALPT model, one could try to think of ways to augment LPT to better represent these non-linear density features.

The virialized motion of FoGs is prohibitively hard to model as a direct function of the signal (which is necessary for the derivative of $P(s)$). It is possible to fix this to some extent by adding a dispersion term to the velocity field model (Heß et al. 2013; Kitaura et al. 2014). However, this implementation needs an extra stochastic step to sample from a Gaussian velocity distribution (the mean and dispersion of which depend on the local density). This breaks the philosophy of BARCODE, in which we sample from one posterior that integrates all the available models. It might be possible to find a way to add a similar model to our posterior. This is a topic of ongoing work.

Another possible way would be to split up very high-density particles into a distribution of smaller particles with a virialized velocity distribution. An easy way to identify these particles would be to use the SC criterion of the $\vartheta = -3$ lower limit. To simplify matters, one could collect clustered particles and collectively split them into one big virialized kernel. However, this would then destroy any remaining substructure and shape.

The triple value regions may require a similar augmentation as the virialized motion, but may be a simpler form is sufficient. It is possible that in ALPT these regions are already partly modelled. However, whether the velocities are also correct in ALPT should be further investigated. If so, this would indeed solve the triple value region problem.

5.2.7 Theoretical considerations

In the relativistic limit, the correspondence between our theoretical RSS definition and observational redshifts breaks down. It might be worth looking into whether a relativistic RSS mapping could be fashioned. One would ideally also include in this the proper cosmological distance instead of the Hubble approximation that we usually make. Such a mapping may lead to rather small corrections on small scales. However, on the largest possible scales, quadratic and higher order terms must be taken into account, as they become dominant. Furthermore, one should start thinking about relativistic structure formation equations.

Within the limits of the framework described in this work, the above cosmological distance arguments are moot. A far more important consideration would be that at some point the data that are used can no longer be assumed to be in the same evolutionary state. Conceivably, LPT should be adaptable to this. The calculation of a field's evolved state at one point in time is just as easily done as another point in time. One could then imagine that each particle in the field is simply evolved to a different point in time, depending on where it ends up. The MCMC nature of the model should be able to iteratively find optimal values for this.

In Section 2, we derived the RSS equations necessary for an application in a fully non-plane parallel RSS. Subsequently, we neglected these terms, due to our plane parallel approximation. We

have not implemented nor tested the additional terms that the full treatment entails. However, this can easily be done. The algorithm will then become fractionally slower, but not significantly. For real galaxy redshift catalogues, the plane parallel approximation cannot be used. When one wants to use these to reconstruct real Universe densities, one must implement the full RSS equations.

6 CONCLUSIONS

Galaxy surveys provide information of the luminous matter distribution, such as galaxy, or cluster catalogues. They are affected by selection effects, yielding an incomplete biased and noisy sample of the underlying non-linear matter distribution. In addition, the tracers are affected by their peculiar motions causing RSDs. A proper analysis of the cosmological large-scale structure should account for all these effects. To this end, Bayesian methods have been developed in the past years connecting the observed data to the primordial density fluctuations, which summarize all the cosmic information for a given set of cosmological parameters and structure formation model.

In this work, we have presented BARCODE, an algorithm for the analysis of the large-scale structure, which for the first time self-consistently solves for coherent RSDs within an analytical Bayesian framework. The contribution of this work in the context of Bayesian techniques relies on the analytical derivation of the Hamiltonian equations taking into account the transformation of the biased tracer to RSS. We present its numerical implementation and a number of tests.

This method could be extended to deal with virialized motions by including a random term accounting for it in the equations, or by a prior FoGs collapse, as suggested in a number of works.

Here, we have chosen the grid-based Bayesian approach, which can easily account for survey mask and geometry of a typical galaxy survey. We have restricted the numerical study to the ZA, but have made derivations including the tidal field tensor within second-order LPT and small-scale SC-based corrections. This approach could also be extended at the expense of a higher computational cost to particle mesh solvers.

From the detailed analytical derivations and numerical tests here presented, we draw the following main conclusions:

- (i) Using our self-contained RSS model, we can overcome large-scale RSD effects in observations and reconstruct the true densities to a great degree of accuracy.
- (ii) When our model is not applied, but rather a naive model based purely on Eulerian real space is used, the RSDs create an anisotropic imprint on the real space reconstructions (Appendix C). We propose to call this effect the *Kaiser effect echo* or Kaiser echo for short.
- (iii) We have demonstrated that BARCODE yields unbiased initial density fluctuations from a biased tracer of the dark matter density field in RSS in terms of the cell-to-cell correlation of the true and reconstructed density field, the respective power spectra and the 2D correlation functions.
- (iv) In particular, we find that the features of the power spectra characterized by the cosmic variance of the considered volume are recovered in detail solving for the Kaiser factor and beyond (as coherent flows have also an impact on non-linear RSDs), and that the 2D correlation functions are isotropized after applying our method.
- (v) Component-wise scrutiny of the Hamiltonian likelihood force in RSS (Appendix B) shows that it contains an unexpected non-

radial term that points in the direction of the displacement field and has an amplitude that grows inversely with distance to the observer. This component may play a strong role when the plane parallel approximation is abandoned. This remains to be investigated.

(vi) Our adaptive leap-frog time-step method was shown to give good MCMC chain performance, yielding uncorrelated samples with a high, stable degree of structural variability, showing that the chain successfully explores the posterior parameter space of initial conditions that match the input (mock) data.

(vii) The number of leap-frog steps per iteration of 256 is sufficient for good chain performance and may even be reduced slightly for this set of parameters (box size, resolution, etc.).

(viii) The RSS model constrains the sampler less strongly, which means the chain has to progress through the posterior space slightly slower than a regular space model. Our adaptive ϵ scheme automatically takes care of this given a large enough N_ϵ value. When N_ϵ is too small (as in Bos (2016)) subsequent samples will be more strongly correlated and the sampler will cover a smaller part of the parameter space.

This code can find applications in a large variety of cosmological problems, such as BAOs reconstruction or cosmic web reconstruction from the $\text{Ly}\alpha$ forest, galaxy, or cluster distributions. The here presented code, which is made publicly available,⁴ thus has the flexibility to tackle realistic large-scale structure analysis from galaxy cluster survey data.

ACKNOWLEDGEMENTS

We are grateful for the opportunity offered by IAU Symposium 308 ‘the Zeldovich Universe’ (June 2014) to present the details of the code and formalism described in this paper.⁵ We express our gratitude to the anonymous reviewer for their critical reading of the initial manuscript and their many incisive questions and suggestions, which no doubt have improved this paper significantly. EGPB and RvdW thank the Leibniz-Institut für Astrophysik Potsdam for the great hospitality during the various work visits that defined the basis for this publication. FSK thanks the Karl-Schwarzschild-Fellowship funding for allowing him to regularly invite EGPB to Potsdam in this period. EGPB further thanks his eScience Center managers Jisk Attema and Rob van Nieuwpoort for supporting the final editing of this manuscript during the past four years. We thank Jelle Kaastra for the encouragement and motivation given during (mainly the early, defining stages of) this project, and Johan Hidding, Maarten Breddels, and Bernard Jones for many interesting and encouraging discussions. EGPB acknowledges support by the NOVA project 10.1.3.07. We would like to thank the Center for Information Technology of the University of Groningen for their support and for providing access to the Peregrine high performance computing cluster. For the analysis and the figures created for this paper, we used a number of scientific software libraries, most importantly: Matplotlib (Hunter 2007) and Seaborn (Waskom et al. 2018) for visualization, NumPy and SciPy (Jones et al. 2018) for numerical computation, and IPython (Pérez & Granger 2007) and Jupyter (Kluyver et al. 2016) for interactive analysis. BARCODE itself was written in C++, using the FFTW3 (Frigo & Johnson 2005) and GSL (Galassi et al. 2009) libraries.

⁴Our source code under MIT license can be found at <https://ascl.net/1810.002> and <https://github.com/egpbos/barcode>.

⁵The talk can be found at <https://doi.org/10.5281/zenodo.1451723>.

REFERENCES

- Alpaslan M. et al., 2014, *MNRAS*, 440, L106
- Ata M., Kitaura F.-S., Müller V., 2015, *MNRAS*, 446, 4250
- Ata M. et al., 2017, *MNRAS*, 467, 3993
- Bel J. et al., 2014, *A&A*, 563, A37
- Berlind A. A., Narayanan V. K., Weinberg D. H., 2001, *ApJ*, 549, 688
- Bernardeau F., Colombi S., Gaztañaga E., Scoccimarro R., 2002, *Phys. Rep.*, 367, 1
- Bertschinger E., 1987, *ApJ*, 323, L103
- Beutler F. et al., 2014, *MNRAS*, 443, 1065
- Birkin J., Li B., Cautun M., Shi Y., 2019, *MNRAS*, 483, 5267
- Birkinshaw M., 1999, *Phys. Rep.*, 310, 97
- Blake C. et al., 2011, *MNRAS*, 415, 2876
- Blake C. et al., 2013, *MNRAS*, 436, 3089
- Bleem L. E. et al., 2015, *ApJS*, 216, 27
- Böhringer H. et al., 2000, *ApJS*, 129, 435
- Bond J. R., Kofman L., Pogosyan D., 1996, *Nature*, 380, 603
- Bos E. G. P., 2016, PhD thesis, Univ. Groningen
- Bos E. G. P., van de Weygaert R., Dolag K., Pettorino V., 2012, *MNRAS*, 426, 440
- Bos E. G. P., van de Weygaert R., Kitaura F., Cautun M., 2016, in van de Weygaert R., Shandarin S., Saar E., Einasto J., eds, *Proc. IAU Symp. Vol. 308, The Zeldovich Universe: Genesis and Growth of the Cosmic Web*. Cambridge University Press, Cambridge, UK, p. 271
- Branchini E., Eldar A., Nusser A., 2002, *MNRAS*, 335, 53
- Brenier Y., Frisch U., Hénon M., Loeper G., Matarrese S., Mohayaee R., Sobolevskii A., 2003, *MNRAS*, 346, 501
- Cautun M., van de Weygaert R., Jones B. J. T., Frenk C. S., 2014, *MNRAS*, 441, 2923
- Cautun M. C., van de Weygaert R., 2011, *Astrophysics Source Code Library*, record on ascl:1105.003
- Challinor A., Lewis A., 2011, *Phys. Rev. D*, 84, 043516
- Cole S., Fisher K. B., Weinberg D. H., 1995, *MNRAS*, 275, 515
- Connor T. et al., 2018, *ApJ*, 867, 25
- da Ângela J. et al., 2008, *MNRAS*, 383, 565
- de Jong J. T. A., Verdoes Kleijn G., Kuijken K., Valentijn E. A., for the KiDS Collaboration, *Astro-WISE Consortiums*, 2012, *Exp. Astron.*, 34
- de la Torre S. et al., 2013, *A&A*, 557, A54
- Desjacques V., Jeong D., Schmidt F., 2018, *Phys. Rep.*, 733, 1
- Doumler T., Hoffman Y., Courtois H., Gottlöber S., 2013, *MNRAS*, 430, 888
- Eckert D., Molendi S., Gastaldello F., Rossetti M., 2011, *A&A*, 529, A133
- Eisenstein D. J., Seo H.-J., Sirko E., Spergel D. N., 2007, *ApJ*, 664, 675
- Fisher K. B., Lahav O., Hoffman Y., Lynden-Bell D., Zaroubi S., 1995, *MNRAS*, 272, 885
- Frigo M., Johnson S. G., 2005, *Proc. IEEE*, 93, 216
- Galassi M. et al., 2009, *GNU Scientific Library Reference Manual*, 3rd edn. Network Theory Ltd, UK
- Gottloeber S., Hoffman Y., Yepes G., 2010, in Wagner S., Steinmetz M., Bode A., Müller M. M., eds, *High Performance Computing in Science and Engineering*, Garching/Munich 2009. Springer, Berlin Heidelberg, p. 309
- Gramann M., 1993, *ApJ*, 405, 449
- Granett B. R. et al., 2015, *A&A*, 583, A61
- Guzzo L. et al., 2008, *Nature*, 451, 541
- Hamilton A. J. S., 1998, in Hamilton D., ed., *Astrophysics and Space Science Library*, Vol. 231, *The Evolving Universe*. Kluwer Academic, Dordrecht, p. 185
- Hawkins E. et al., 2003, *MNRAS*, 346, 78
- Heß S., Kitaura F.-S., Gottlöber S., 2013, *MNRAS*, 435, 2065
- Hidding J., van de Weygaert R., Shandarin S., 2016, in van de Weygaert R., Shandarin S., Saar E., Einasto J., eds, *Proc. IAU Symp. 308, The Zeldovich Universe: Genesis and Growth of the Cosmic Web*. Cambridge University Press, Cambridge, UK, p. 69
- Hilton M. et al., 2018, *ApJS*, 235, 20
- Hoekstra H., Bartelmann M., Dahle H., Israel H., Limousin M., Meneghetti M., 2013, *Space Sci. Rev.*, 177, 75
- Hoffman M. D., Gelman A., 2014, *J. Mach. Learn. Res.*, 15, 1593
- Hoffman Y., Ribak E., 1991, *ApJ*, 380, L5
- Hunter J. D., 2007, *Comput. Sci. Eng.*, 9, 90
- Icke V., 1973, *A&A*, 27, 1
- Ikebe Y. et al., 1996, *Nature*, 379, 427
- Jackson J. C., 1972, *MNRAS*, 156, 1P
- Jain B., Zhang P., 2008, *Phys. Rev. D*, 78, 063503
- Jasche J., Kitaura F. S., 2010, *MNRAS*, 407, 29
- Jasche J., Lavaux G., 2018, *A&A*, 625, A64
- Jasche J., Wandelt B. D., 2013, *MNRAS*, 432, 894
- Jaynes E., Bretthorst G., 2003, *Probability Theory: The Logic of Science*. Cambridge Univ. Press, Cambridge
- Jennings E., Baugh C. M., Pascoli S., 2011, *MNRAS*, 410, 2081
- Jones E. et al., 2001–2018, *SciPy: Open source scientific tools for Python*. <http://www.scipy.org/>
- Kaiser N., 1984, *ApJ*, 284, L9
- Kaiser N., 1987, *MNRAS*, 227, 1
- Kaiser N., Squires G., 1993, *ApJ*, 404, 441
- King I. R., 1972, *ApJ*, 174, L123
- Kitaura F.-S., 2013, *MNRAS*, 429, L84
- Kitaura F.-S., Angulo R. E., 2012, *MNRAS*, 425, 2443
- Kitaura F. S., Enßlin T. A., 2008, *MNRAS*, 389, 497
- Kitaura F.-S., Heß S., 2013, *MNRAS*, 435, L78
- Kitaura F.-S., Erdoğan P., Nuza S. E., Khalatyan A., Angulo R. E., Hoffman Y., Gottlöber S., 2012, *MNRAS*, 427, L35
- Kitaura F.-S., Yepes G., Prada F., 2014, *MNRAS*, 439, L21
- Kitaura F.-S., Gil-Marín H., Scóccola C. G., Chuang C.-H., Müller V., Yepes G., Prada F., 2015, *MNRAS*, 450, 1836
- Kitaura F.-S. et al., 2016a, *MNRAS*, 456, 1456
- Kitaura F.-S., Ata M., Angulo R. E., Chuang C.-H., Rodríguez-Torres S., Monteagudo C. H., Prada F., Yepes G., 2016b, *MNRAS*, 457, L113
- Kluyver T. et al., 2016, in Loizides F., Schmidt B., eds, *Positioning and Power in Academic Publishing: Players, Agents and Agendas*. IOS Press, Amsterdam, p. 87
- Komatsu E. et al., 2011, *ApJS*, 192, 18
- Kwan J., Lewis G. F., Linder E. V., 2012, *ApJ*, 748, 78
- Lagattuta D. J. et al., 2019, *MNRAS*, 485, 3738
- Laureijs R. et al., 2011, preprint ([arXiv:1110.3193](https://arxiv.org/abs/1110.3193))
- Lavaux G., Wandelt B. D., 2010, *MNRAS*, 403, 1392
- Leclercq F., 2015, PhD thesis, Institut d'Astrophysique de Paris ([arXiv:1512.04985](https://arxiv.org/abs/1512.04985))
- Leclercq F., Jasche J., Wandelt B., 2015a, *J. Cosmol. Astropart. Phys.*, 6, 15
- Leclercq F., Jasche J., Wandelt B., 2015b, *A&A*, 576, L17
- Leclercq F., Jasche J., Gil-Marín H., Wandelt B., 2013, *J. Cosmol. Astropart. Phys.*, 11, 48
- Lee K.-G. et al., 2016, *ApJ*, 817, 160
- Lee M. H., Babul A., Kofman L., Kaiser N., 1997, *ApJ*, 489, 522
- McDonald P., Seljak U., 2009, *J. Cosmol. Astropart. Phys.*, 10, 007
- Mohayaee R., Mathis H., Colombi S., Silk J., 2006, *MNRAS*, 365, 939
- Monaco P., Efstathiou G., 1999, *MNRAS*, 308, 763
- Monaco P., Sefusatti E., Borgani S., Crocce M., Fosalba P., Sheth R. K., Theuns T., 2013, *MNRAS*, 433, 2389
- Monaghan J. J., 1992, *ARA&A*, 30, 543
- Mulchaey J. S., 2000, *ARA&A*, 38, 289
- Neal R. M., 1993, *CRG-TR-93-1 Technical Report*, Probabilistic inference using Markov chain Monte Carlo methods. Univ. Toronto, Toronto, p. 144
- Neal R. M., 2012, preprint ([arXiv:1206.1901](https://arxiv.org/abs/1206.1901))
- Nesseris S., Perivolaropoulos L., 2008, *Phys. Rev. D*, 77, 023504
- Neyrinck M. C., 2013, *MNRAS*, 428, 141
- Neyrinck M. C., 2016, *MNRAS*, 455, L11
- Neyrinck M. C., Aragón-Calvo M. A., Jeong D., Wang X., 2014, *MNRAS*, 441, 646
- Nulsen P. E. J., Powell S. L., Vikhlinin A., 2010, *ApJ*, 722, 55
- Nusser A., Branchini E., 2000, *MNRAS*, 313, 587
- Nusser A., Davis M., 1994, *ApJ*, 421, L1
- Nusser A., Dekel A., 1992, *ApJ*, 391, 443
- Okumura T., Seljak U., Desjacques V., 2012, *J. Cosmol. Astropart. Phys.*, 11, 014

- Okumura T., Seljak U., Vlah Z., Desjacques V., 2014, *J. Cosmol. Astropart. Phys.*, 5, 003
- Okumura T., Matsumura T., Eisenstein D. J., Kayo I., Hikage C., Szalay A. S., Schneider D. P., 2008, *ApJ*, 676, 889
- Park D., Lee J., 2007, *Phys. Rev. Lett.*, 98, 081301
- Peacock J. A. et al., 2001, *Nature*, 410, 169
- Peebles P. J. E., 1989, *ApJ*, 344, L53
- Percival W. J., White M., 2009, *MNRAS*, 393, 297
- Percival W. J. et al., 2004, *MNRAS*, 353, 1201
- Pérez F., Granger B. E., 2007, *Comput. Sci. Eng.*, 9, 21
- Pizzo R. F., de Bruyn A. G., Bernardi G., Brentjens M. A., 2011, *A&A*, 525, A104
- Planck Collaboration XXIV, 2016, *A&A*, 594, A24
- Planck Collaboration X, 2018a, preprint ([arXiv:1807.06211](https://arxiv.org/abs/1807.06211))
- Planck Collaboration VI, 2018b, preprint ([arXiv:1807.06209](https://arxiv.org/abs/1807.06209))
- Radovich M. et al., 2017, *A&A*, 598, A107
- Reid B. A. et al., 2012, *MNRAS*, 426, 2719
- Ryden B. S., Melott A. L., 1996, *ApJ*, 470, 160
- Samsing J., Skielboe A., Hansen S. H., 2012, *ApJ*, 748, 21
- Samushia L., Percival W. J., Raccanelli A., 2012, *MNRAS*, 420, 2102
- Samushia L. et al., 2013, *MNRAS*, 429, 1514
- Samushia L. et al., 2014, *MNRAS*, 439, 3504
- Sánchez A. G. et al., 2014, *MNRAS*, 440, 2692
- Sarazin C. L., 1986, *Rev. Mod. Phys.*, 58, 1
- Sargent W. L. W., Turner E. L., 1977, *ApJ*, 212, L3
- Sartoris B. et al., 2016, *MNRAS*, 459, 1764
- Schaap W. E., van de Weygaert R., 2000, *A&A*, 363, L29
- Schaller M. et al., 2015, *MNRAS*, 451, 1247
- Schindler S., 1996, *A&A*, 305, 756
- Schmittfull M., Feng Y., Beutler F., Sherwin B., Chu M. Y., 2015, *Phys. Rev. D*, 92, 123522
- Song Y.-S., Koyama K., 2009, *J. Cosmol. Astropart. Phys.*, 1, 048
- Song Y.-S., Percival W. J., 2009, *J. Cosmol. Astropart. Phys.*, 10, 004
- Song Y.-S., Sabiu C. G., Nichol R. C., Miller C. J., 2010, *J. Cosmol. Astropart. Phys.*, 1, 025
- Song Y.-S., Sabiu C. G., Kayo I., Nichol R. C., 2011, *J. Cosmol. Astropart. Phys.*, 5, 020
- Sorce J. G., Tempel E., 2017, *MNRAS*, 469, 2859
- Sorce J. G., Tempel E., 2018, *MNRAS*, 476, 4362
- Springel V., 2010, *ARA&A*, 48, 391
- Starobinsky A. A., 1980, *Phys. Lett. B*, 91, 99
- Tanimura H., Aghanim N., Douspis M., Beelen A., Bonjean V., 2019, *A&A*, 625, A67
- Tassev S., Zaldarriaga M., Eisenstein D. J., 2013, *J. Cosmol. Astropart. Phys.*, 6, 36
- Tassev S., Eisenstein D. J., Wandelt B. D., Zaldarriaga M., 2015, preprint ([arXiv:1502.07751](https://arxiv.org/abs/1502.07751))
- Taylor J. F., Ashdown M. A. J., Hobson M. P., 2008, *MNRAS*, 389, 1284
- Tegmark M. et al., 2004, *ApJ*, 606, 702
- Tempel E., Tago E., Liivamägi L. J., 2012, *A&A*, 540, A106
- Tempel E., Kipper R., Tamm A., Gramann M., Einasto M., Sepp T., Tuvikene T., 2016, *A&A*, 588, A14
- The LSST Dark Energy Science Collaboration, 2018, preprint ([arXiv:1809.01669](https://arxiv.org/abs/1809.01669))
- Thomas B. C., Melott A. L., Feldman H. A., Shandarin S. F., 2004, *ApJ*, 601, 28
- Tojeiro R. et al., 2014, *MNRAS*, 440, 2222
- Tully R. B., 2015, *AJ*, 149, 171
- Tully R. B., Fisher J. R., 1978, in Longair M. S., Einasto J., eds, *Proc. IAU Symp. 79, Large-scale structures in the Universe*. Reidel, Dordrecht, p. 31
- Umetsu K., 2010, preprint ([arXiv:1002.3952](https://arxiv.org/abs/1002.3952))
- Vakili M., Kitaura F.-S., Feng Y., Yepes G., Zhao C., Chuang C.-H., Hahn C., 2017, *MNRAS*, 472, 4144
- van de Weygaert R., 1996, in Coles P., Martinez V., Pons-Borderia M.-J., eds, *ASP Conf. Ser. Vol. 94, Mapping, Measuring, and Modelling the Universe*. Astron. Soc. Pac., San Francisco, p. 49
- van de Weygaert R., Bertschinger E., 1996, *MNRAS*, 281, 84
- van de Weygaert R., Platen E., 2011, *Int. J. Mod. Phys. Conf. Ser.*, 1, 41
- van Haarlem M., van de Weygaert R., 1993, *ApJ*, 418, 544
- Vogelsberger M. et al., 2014, *MNRAS*, 444, 1518
- Wang H., Mo H. J., Yang X., van den Bosch F. C., 2012, *MNRAS*, 420, 1809
- Wang H., Mo H. J., Yang X., van den Bosch F. C., 2013, *ApJ*, 772, 63
- Wang H., Mo H. J., Yang X., Jing Y. P., Lin W. P., 2014, *ApJ*, 794, 94
- Wang H., Mo H. J., Jing Y. P., Guo Y., van den Bosch F. C., Yang X., 2009, *MNRAS*, 394, 398
- Waskom M. et al., 2018, mwaskom/seaborn: v0.9.0 (July 2018). <https://doi.org/10.5281/zenodo.1313201>
- Weinberg D. H., 1992, *MNRAS*, 254, 315
- Werner N., Finoguenov A., Kaastra J. S., Simionescu A., Dietrich J. P., Vink J., Böhringer H., 2008, *A&A*, 482, L29
- White M., 2015, *MNRAS*, 450, 3822
- White M., Song Y.-S., Percival W. J., 2009, *MNRAS*, 397, 1348
- White M., Carlson J., Reid B., 2012, *Redshift Space Distortions and The growth of Cosmic Structure*. <http://mwhite.berkeley.edu/Talks/SantaFe12.BOSS.pdf>
- Yahil A., Strauss M. A., Davis M., Huchra J. P., 1991, *ApJ*, 372, 380
- Zaroubi S., Hoffman Y., Fisher K. B., Lahav O., 1995, *ApJ*, 449, 446
- Zel'dovich Y. B., 1970, *A&A*, 5, 84
- Zhang P., Liguori M., Bean R., Dodelson S., 2007, *Phys. Rev. Lett.*, 99, 141302
- Zhao G.-B., Giannantonio T., Pogosian L., Silvestri A., Bacon D. J., Koyama K., Nichol R. C., Song Y.-S., 2010, *Phys. Rev. D*, 81, 103510
- Zheng Y., Song Y.-S., 2016, *J. Cosmol. Astropart. Phys.*, 2016, 050
- Zheng Y., Zhang P., Jing Y., Lin W., Pan J., 2013, *Phys. Rev. D*, 88, 103510

APPENDIX A: BARCODE: HAMILTONIAN MONTE CARLO SAMPLING ALGORITHM

In this section, we briefly recap the basic concepts, quantities, and formulas behind BARCODE that are necessary for this paper. See Section 2.1 for a general overview of the components described in more detail in this appendix. Since the HMC algorithm, specifically in cosmology, has also been explained in great detail in previous works (Neal 1993; Jasche & Kitaura 2010; Jasche & Wandelt 2013; Wang et al. 2013; Bos 2016), we keep it short here.

A1 Hamiltonian Monte Carlo terminology

The Hamiltonian \mathcal{H} is the sum of the potential energy \mathcal{E} and the kinetic energy \mathcal{K} :

$$\mathcal{H} = \mathcal{E} + \mathcal{K}. \quad (\text{A1})$$

The kinetic term in this is the momentum (row-vector) times the inverse of the mass matrix \mathcal{M} times the momentum (column-vector):

$$\mathcal{K} = \frac{1}{2} \dot{\mathbf{p}}^T \mathcal{M}^{-1} \dot{\mathbf{p}}. \quad (\text{A2})$$

The potential \mathcal{E} is coupled to the PDF of our signal $\delta(\mathbf{q}_i)$. We define this PDF in a Bayesian way as the posterior P , which is the product of a prior \mathcal{P} with a likelihood \mathcal{L} .⁶ The connection between the potential and the PDF is given as a canonical ensemble distribution. In logarithmic form, this gives us an equation for \mathcal{E} in terms of the posterior P :

$$\mathcal{E} = \text{constant} - \ln P = \text{constant} - \ln \mathcal{P} - \ln \mathcal{L}, \quad (\text{A3})$$

⁶Note that the notation we use is not necessarily the most common in Bayesian literature. For instance, Jaynes & Bretthorst (2003, page 89) uses a notation that would translate in this work to something like $L(\delta(\mathbf{q})) = P(\delta^{\text{obs}}(\mathbf{x})|\delta(\mathbf{q}))$, where L is the ‘likelihood’ of the hypothesis and P is a probability function that describes it in terms of the random variable $\delta^{\text{obs}}(\mathbf{x})$ and conditional variable $\delta(\mathbf{q})$ (the hypothesis).

so $-\ln \mathcal{P}$ and $-\ln \mathcal{L}$ are the terms we need our (statistical) model to specify so that we can calculate \mathcal{E} , up to a constant. However, as we will see next, we only need to know $\delta \mathcal{H}$ or the derivative of the potential. We can therefore forget about the constants.

In this work, we use a Gaussian prior and likelihood:

$$-\ln \mathcal{P} = \text{constant} + \frac{1}{2} \sum_{i,j} \delta(\mathbf{q}_i) (S^{-1})_{ij} \delta(\mathbf{q}_j), \quad (\text{A4})$$

$$-\ln \mathcal{L} = \text{constant} + \frac{1}{2} \sum_i \frac{[T(\rho^{\text{obs}}(\mathbf{x}_i)) - \rho(\mathbf{x}_i)]^2 w_i}{\sigma_i^2}. \quad (\text{A5})$$

These are two different types of Gaussian exponents. The likelihood one is a simple sum of 1D Gaussians; one for each grid cell. The prior exponent is a little more computationally intensive, due to the matrix inversion. However, we can approximate the $(S^{-1})_{ij} \delta(\mathbf{q}_j)$ by treating it as a convolution in real space and thus a simple multiplication in Fourier space, where the (inverse) power spectrum (the FT of S^{-1}) is known and easy (diagonal) (see appendix 3.B.3 of Bos 2016 for computational details). The same considerations apply for the kinetic term \mathcal{K} , though the mass matrix \mathcal{M} is a different quantity. Often, it is taken to be $\mathcal{M} = S^{-1}$, meaning that indeed a similar procedure can be applied, only now with the non-inverted power spectrum.

A2 Leap frog scheme

After drawing a random momentum vector, we ‘move’ the MCMC chain forward through the posterior’s parameter space by solving the Hamiltonian equations of motion using a leap-frog estimation scheme. Numerically, we will be solving the following discretized equations for each time-step $\tau + \epsilon$ (based on the previous step at τ):

$$\begin{aligned} \check{p}_i \left(\tau + \frac{\epsilon}{2} \right) &\simeq \check{p}_i(\tau) - \frac{\epsilon}{2} \frac{\partial \mathcal{E}}{\partial \check{q}_i}(\check{q}(\tau)) \\ \check{q}_i(\tau + \epsilon) &\simeq \check{q}_i(\tau) + \epsilon \sum_j (\mathcal{M}^{-1})_{ij} \check{p}_j \left(\tau + \frac{\epsilon}{2} \right) \end{aligned} \quad (\text{A6})$$

$$\check{p}_i(\tau + \epsilon) \simeq \check{p}_i \left(\tau + \frac{\epsilon}{2} \right) - \frac{\epsilon}{2} \frac{\partial \mathcal{E}}{\partial \check{q}_i}(\check{q}(\tau + \epsilon)).$$

We will need the derivative of \mathcal{E} to each component of the signal, i.e. every $\delta(\mathbf{q}_i)$ sampled on a regular grid $\{\mathbf{q}_i\}$:

$$F_i \equiv \frac{\partial \mathcal{E}}{\partial \delta(\mathbf{q}_i)} = F_i^{\mathcal{P}} + F_i^{\mathcal{L}}. \quad (\text{A7})$$

This F_i is called the Hamiltonian force and its components are called the prior force and likelihood force, respectively. For our Gaussian prior, the prior force is

$$F_i^{\mathcal{P}} = \sum_j (S^{-1})_{ij} \delta(\mathbf{q}_j), \quad (\text{A8})$$

where S_{ij} is the correlation matrix corresponding to the cosmological power spectrum, given an inverse power spectrum mass (Section 3.2).

In general, we can write the likelihood force as

$$F_k^{\mathcal{L}} = \sum_m h(\mathbf{q}_m) \frac{\partial \vartheta(\mathbf{q}_m)}{\partial \delta(\mathbf{q}_k)}, \quad (\text{A9})$$

where the derivative of the divergence of the displacement field $\vartheta(\mathbf{q})$ to the signal $\delta(\mathbf{q}_k)$ encodes the structure formation model and

the h term primarily represents the Gaussian model that we use for our likelihood (see chapter 3 of Bos 2016 for more details). For the Zel’dovich model of structure formation, the likelihood force is given by

$$F_i^{\mathcal{L}, \text{ZA}} = -h(\mathbf{q}_i). \quad (\text{A10})$$

The function $h(\mathbf{q})$ is subsequently defined as

$$\hat{h}(\mathbf{k}_l) = - \left(-\frac{i\mathbf{k}_l}{k_l^2} \cdot (\hat{\mathbf{V}}'(\mathbf{k}_l))^* \right)^* = -\frac{i\mathbf{k}_l}{k_l^2} \cdot \hat{\mathbf{V}}_l, \quad (\text{A11})$$

or

$$h(\mathbf{q}_m) = \nabla^{-2} \nabla \cdot \mathbf{V}_m, \quad (\text{A12})$$

where

$$\begin{aligned} \mathbf{V}_i &\equiv \mathbf{V}(\mathbf{x}(\mathbf{q}_i)) \equiv \frac{\partial \ln \mathcal{L}}{\partial \mathbf{x}(\mathbf{q}_i)} \\ &= m_i \sum_j \frac{\partial \ln \mathcal{L}}{\partial \rho(\mathbf{x}_j)} \frac{\partial W(\mathbf{x}_j - \mathbf{x}(\mathbf{q}_i); h_s)}{\partial \mathbf{x}(\mathbf{q}_i)}. \end{aligned} \quad (\text{A13})$$

Here, the derivative of the likelihood to the density in Eulerian space $\rho(\mathbf{x})$ for a Gaussian likelihood is given by

$$\frac{\partial \ln \mathcal{L}}{\partial \rho(\mathbf{x}_j)} = \frac{(T_j(\rho^{\text{obs}}(\mathbf{x}_j)) - \rho(\mathbf{x}_j)) w_j}{\sigma_j^2}. \quad (\text{A14})$$

Note that at this point, a mixing of the Lagrangian and Eulerian coordinate systems has been introduced. The signal grid denoted here with subscript i corresponds to the Lagrangian space, which of the primordial density field that we are interested in sampling from our PDF. The grid denoted by subscript j , in turn, corresponds to the discretized Eulerian grid of the observations. This j index is summed out. Hence, we are left in the end with the Lagrangian grid cells only. It is especially recommended to keep this distinction in mind when considering the derivations in Appendix B.

A2.1 Leap frog time-step precision

The maximum pseudo time-step ϵ_m is probably the most important parameter that determines overall HMC performance. The pseudo time-step ϵ is drawn randomly from a uniform distribution between 0 and ϵ_m .

If ϵ_m is too high, the leap-frog errors will be too large, leading to artificial increase in energy. The acceptance criterion does not allow this, meaning that many steps will be rejected. This wastes computational resources.

On the other hand, if it is too low, many steps will be accepted. However, the samples will be very close together, meaning that the probability distribution is densely and slowly sampled. This is an inefficient use of computational power since we do not need a too dense sampling, but rather a representative sampling of the overall high-probability regions.

Striking a balance between the two is essential. Further complications arise from the fact that the energy also scales with the dimensionality of the problem N . To keep the acceptance rate at an acceptable value, ϵ will have to be smaller for larger problems, meaning that the chain will evolve correspondingly slower. It is therefore impossible to determine a universally optimal ϵ for every HMC problem. Neal (2012) discusses the ϵ issue to great lengths.

One of the possible solutions is to automatically adapt ϵ_m to get an optimal acceptance rate (Hoffman & Gelman 2012; Neal 2012). We implemented our own adaptive scheme in BARCODE. In accordance with Neal (2012), we set the target acceptance rate to

0.65. The code will check after a set number of attempted iterations (50 in our case) if the acceptance rate is still within a reasonable range around our optimal rate (± 0.05 is allowed). If it is not, ϵ_m is adjusted an appropriate amount in the right direction. Furthermore, to accelerate the initial adaptive stage (the first few HMC iterations) the code checks for acceptance every step instead of checking the acceptance rate every 50 steps.

A3 Hybrid Monte Carlo acceptance criterion

Like Metropolis–Hastings and Gibbs sampling, Hamiltonian sampling includes a sample rejection step after the candidate sampling step. Given the difference $\delta\mathcal{H}$ between the Hamiltonian of the state (\tilde{q}, \tilde{p}) (where $\tilde{q} = \delta(q)$) before the dynamical step, $\mathcal{H}(\tilde{q}, \tilde{p})$, and that of the candidate state $(\check{q}^*, \check{p}^*)$ after dynamics, $\mathcal{H}(\check{q}^*, \check{p}^*)$, the probability of acceptance of the candidate state conditional on the previous state $P((\check{q}^*, \check{p}^*) | (\tilde{q}, \tilde{p}))$ is given by

$$P((\check{q}^*, \check{p}^*) | (\tilde{q}, \tilde{p})) = \min \left(1, e^{-(\mathcal{H}(\check{q}^*, \check{p}^*) - \mathcal{H}(\tilde{q}, \tilde{p}))} \right). \quad (\text{A15})$$

A4 From Lagrangian to Eulerian coordinates

A final important aspect of BARCODE – which is unrelated to HMC, but crucial to our cosmological application – concerns the transformation between the Lagrangian space of the sampled primordial density fields and that of the Eulerian comoving coordinates at some $a > 0$. Lagrangian coordinates q can, in general, be transformed to their Eulerian counterparts x by adding a displacement field Ψ (in our case, defined by the ZA):

$$x(q) = q + \Psi(q). \quad (\text{A16})$$

Note that this relation cannot be inverted. Mathematically, this is because the displacement field is dependent on q . Physically, it is because multiple fluid elements originally at $\{q_i\}$ might end up at one same Eulerian location x ; through gravitational clustering, this will in fact happen all over the place. This means that we must take the q_i grid as the basis of our analysis, as taking an x grid as the basis would make it impossible to find unique corresponding q values, which computationally is a problem. To compute Ψ , we need ϑ because in Fourier space we have

$$ik\tilde{\vartheta} = -k^2\tilde{\Psi} \quad (\text{A17})$$

$$\Rightarrow \tilde{\Psi} = -\frac{ik}{k^2}\tilde{\vartheta}. \quad (\text{A18})$$

For the Zel’dovich model of structure formation that we use in this work, this quantity is given by

$$\vartheta = -D_1\Delta^{(1)}(q). \quad (\text{A19})$$

APPENDIX B: HAMILTONIAN FORCE IN REDSHIFT SPACE

The Hamiltonian likelihood force⁷ – defined for observations in comoving coordinates in equation (A9) – is redefined for observations in RSS as

$$F_i^{\mathcal{L}} = - \sum_j \frac{\partial \ln \mathcal{L}}{\partial \rho^s(s_j)} \frac{\partial \rho^s(s_j)}{\partial \tilde{q}_i}, \quad (\text{B1})$$

⁷The Hamiltonian *prior* force is defined in Lagrangian space only. We need not take RSS into account there.

where $\rho^s(s_j)$ is the Eulerian density field in RSS, at grid location j , which is found by evolving the Lagrangian density field $\tilde{q} = \delta(q)$ forward (see Appendix F), and \tilde{q}_i is the Lagrangian density field (the signal) at grid location i . The first right-hand side multiplicative term in equation (B1) is given analogously to equation (A14) by

$$\frac{\partial \ln \mathcal{L}}{\partial \rho^s(s_j)} = \frac{w_j}{\sigma_j^2} [T(\rho^{\text{obs}}(s_j)) - \rho^s(s_j)], \quad (\text{B2})$$

where $T(\rho)$ is a transfer function that performs a simple linear density transformation to values in the used structure formation model given values from an N -body simulation (which should correspond to the observed quantities). For the non-RSD models, we could use functions from Leclercq et al. (2013). However, the fitted functions from that work were not calibrated for RSS densities, so we would need to rederive such functions. We highlight this possibility for future applications (see Section 5.2.1), but without any loss of generality, we can define $T(\rho) = \rho$ here. Since the T term in this equation does not depend on the primordial density field, its exact form is not important in our derivation.

Note that indices i and j run over completely different regular grids. The i index here refers to the initial Lagrangian space grid, while the j index is for the present-time Eulerian grid defined by the observations.

The density estimation of $\rho^s(s_j)$ is done with SPH splines (Appendix F). In this case, we replace x by s in equation (F1):

$$\rho^s(s) = \sum_i m_i W(s - s_i; h_s). \quad (\text{B3})$$

With that we can write the second right-hand side term of equation (B1) as

$$\begin{aligned} \frac{\partial \rho^s(s_j)}{\partial \tilde{q}_i} &= \sum_k m_k \frac{\partial W(\|s_j - s_k\|)}{\partial \tilde{q}_i} \\ &= \sum_k m_k \frac{\partial W(\|s_j - s_k\|)}{\partial s_k} \cdot \frac{\partial s_k}{\partial \tilde{q}_i}, \end{aligned} \quad (\text{B4})$$

where the gradient of W is given in equation (F6; replacing x by s). The second term is where the real difference with the non-RSD method comes in because

$$\frac{\partial s_k}{\partial \tilde{q}_i} = \frac{\partial \Psi(q_k)}{\partial \tilde{q}_i} + \frac{\partial \Psi_s(q_k)}{\partial \tilde{q}_i}, \quad (\text{B5})$$

where, besides the derivative of Ψ , we now have the extra derivative of RSS displacement Ψ_s term to deal with. It is this term that we will derive further in the following subsections. Putting this back into equation (B1), we can rewrite as

$$\begin{aligned} F_i^{\mathcal{L}} &= - \sum_k V_k \cdot \left[\frac{\partial \Psi_k}{\partial \tilde{q}_i} + \frac{\partial \Psi_{s_k}}{\partial \tilde{q}_i} \right] \\ &= \sum_m h_m \frac{\partial \vartheta_m}{\partial \tilde{q}_i} - \sum_k V_k \cdot \frac{\partial \Psi_{s_k}}{\partial \tilde{q}_i} \\ &\equiv F_i^{\mathcal{L}, \text{reg}} + F_i^{\mathcal{L}, \text{rss}}, \end{aligned} \quad (\text{B6})$$

where the first part is the same as the ‘regular’ $F^{\mathcal{L}}$ without RSS from equation (A9) and the second part represents the RSS contribution.

We carry on to derive in detail the multiplicand in this added term, which is the derivative or gradient of the RSS displacement field Ψ_s with respect to the signal (the primordial density field $\delta(q)$). For

convenience, we multiply by Ha , yielding

$$(Ha) \frac{\partial \Psi_s(\mathbf{q}_k)}{\partial \check{q}_i} = \frac{\partial ((\mathbf{v}_k - \mathbf{v}_{\text{obs}}) \cdot \hat{\mathbf{x}}_k) \hat{\mathbf{x}}_k}{\partial \check{q}_i} \\ = \left(\frac{\partial \mathbf{v}_k}{\partial \check{q}_i} \cdot \hat{\mathbf{x}}_k \right) \hat{\mathbf{x}}_k + \left((\mathbf{v}_k - \mathbf{v}_{\text{obs}}) \cdot \frac{\partial \hat{\mathbf{x}}_k}{\partial \check{q}_i} \right) \hat{\mathbf{x}}_k \\ + ((\mathbf{v}_k - \mathbf{v}_{\text{obs}}) \cdot \hat{\mathbf{x}}_k) \frac{\partial \hat{\mathbf{x}}_k}{\partial \check{q}_i}. \quad (\text{B7})$$

We can further expand the gradient of $\hat{\mathbf{x}}_k$:

$$\frac{\partial \hat{\mathbf{x}}_k}{\partial \check{q}_i} = \frac{1}{\|\mathbf{x}_k\|} \frac{\partial \Psi_k}{\partial \check{q}_i} + \mathbf{x}_k \left[-\frac{1}{2} \frac{2x_k \frac{\partial x_k}{\partial \check{q}_i} + 2y_k \frac{\partial y_k}{\partial \check{q}_i} + 2z_k \frac{\partial z_k}{\partial \check{q}_i}}{(x_k^2 + y_k^2 + z_k^2)^{\frac{3}{2}}} \right] \\ = \frac{1}{\|\mathbf{x}_k\|} \frac{\partial \Psi_k}{\partial \check{q}_i} + \mathbf{x}_k \frac{1}{\|\mathbf{x}_k\|^3} \left(\mathbf{x}_k \cdot \frac{\partial \mathbf{x}_k}{\partial \check{q}_i} \right) \\ = \frac{1}{\|\mathbf{x}_k\|} \left[\frac{\partial \Psi_k}{\partial \check{q}_i} - \left(\hat{\mathbf{x}}_k \cdot \frac{\partial \Psi_k}{\partial \check{q}_i} \right) \hat{\mathbf{x}}_k \right]. \quad (\text{B8})$$

This is the most explicit form we can give in the general case. To find a specific solution for our models, we must specify \mathbf{v} . In LPT, given Ψ , we know \mathbf{v} :

$$\mathbf{v} = aHf^{(1)}\Psi^{(1)} \quad \text{LPT (Zel'dovich)}, \quad (\text{B9})$$

$$\mathbf{v} = aH(f^{(1)}\Psi^{(1)} + f^{(2)}\Psi^{(2)}) \quad \text{2LPT}, \quad (\text{B10})$$

For convenience, in what follows we split Ψ_s into two parts:

$$\Psi_s = \frac{1}{Ha} \left(\sum_o \mathbf{v}^{(o)} - \mathbf{v}_{\text{obs}} \right) \cdot \hat{\mathbf{x}} \hat{\mathbf{x}} \\ = \left(\sum_o f^{(o)}\Psi^{(o)} - \frac{\mathbf{v}_{\text{obs}}}{Ha} \right) \cdot \hat{\mathbf{x}} \hat{\mathbf{x}} \equiv \sum_o \Psi_s^{(o)} - \Psi_s^{\text{obs}}, \quad (\text{B11})$$

where o stands for the LPT order of the \mathbf{v} and corresponding Ψ_s terms. This way we can treat the observer's velocity contribution separately from the LPT part(s). In what follows, we work out the observer's velocity part and the Zel'dovich first-order LPT term.

B1 Zel'dovich term

When we insert equation (B9) into equation (B7) and combine with equation (B8), we find for the derivative of $\Psi_s^{(1)}$ that (note that by Ψ we really mean $\Psi^{(1)}$ here)

$$\frac{1}{f^{(1)}} \frac{\partial \Psi_s^{(1)}}{\partial \check{q}_i} = \left(\frac{\partial \Psi_k}{\partial \check{q}_i} \cdot \hat{\mathbf{x}}_k \right) \hat{\mathbf{x}}_k + \frac{1}{\|\mathbf{x}_k\|} \left[\Psi_k \cdot \left(\frac{\partial \Psi_k}{\partial \check{q}_i} - \left(\hat{\mathbf{x}}_k \cdot \frac{\partial \Psi_k}{\partial \check{q}_i} \right) \hat{\mathbf{x}}_k \right) \right] \hat{\mathbf{x}}_k + \frac{1}{\|\mathbf{x}_k\|} (\Psi_k \cdot \hat{\mathbf{x}}_k) \left[\frac{\partial \Psi_k}{\partial \check{q}_i} - \left(\hat{\mathbf{x}}_k \cdot \frac{\partial \Psi_k}{\partial \check{q}_i} \right) \hat{\mathbf{x}}_k \right] \\ = \left(\frac{\partial \Psi_k}{\partial \check{q}_i} \cdot \hat{\mathbf{x}}_k \right) \hat{\mathbf{x}}_k + \frac{1}{\|\mathbf{x}_k\|} \left[\Psi_k \cdot \frac{\partial \Psi_k}{\partial \check{q}_i} \hat{\mathbf{x}}_k - 2(\Psi_k \cdot \hat{\mathbf{x}}_k) \left(\frac{\partial \Psi_k}{\partial \check{q}_i} \cdot \hat{\mathbf{x}}_k \right) \hat{\mathbf{x}}_k + (\Psi_k \cdot \hat{\mathbf{x}}_k) \frac{\partial \Psi_k}{\partial \check{q}_i} \right]. \quad (\text{B12})$$

To facilitate insight into this expression, we introduce a few auxiliary variables:

$$\frac{\partial \Psi_{s_k}^{(1)}}{\partial \check{q}_i} \equiv f^{(1)} \left[\left(A_{ki}^{(1)} + B_{ki}^{(1)} - 2C_k^{(1)} A_{ki}^{(1)} \right) \hat{\mathbf{x}}_k + C_k^{(1)} \frac{\partial \Psi_k^{(1)}}{\partial \check{q}_i} \right], \quad (\text{B13})$$

following the definition:

$$A_{ki}^{(1)} \equiv \frac{\partial \Psi_k^{(1)}}{\partial \check{q}_i} \cdot \hat{\mathbf{x}}_k \quad (\text{B14})$$

$$B_{ki}^{(1)} \equiv \frac{1}{\|\mathbf{x}_k\|} \Psi_k^{(1)} \cdot \frac{\partial \Psi_k^{(1)}}{\partial \check{q}_i} \quad (\text{B15})$$

$$C_k^{(1)} \equiv \frac{1}{\|\mathbf{x}_k\|} \Psi_k^{(1)} \cdot \hat{\mathbf{x}}_k. \quad (\text{B16})$$

The A and B terms have units $h^{-1}\text{Mpc}$ and C is unitless. It is interesting that, while the first three terms represent a contribution in the radial direction, the last term corresponds to a contribution in the direction of the displacement field. Depending on the density field, this may be an arbitrary direction. The C term in front of it contains a division by the distance to the particle, meaning that the closer you look, the stronger this component will be. The same is true for the B term. The derivative of the RSS displacement field is thus quite different from what one might naively expect from regular (radial) RSD effects, especially near the observer. The A term is the most straightforward, simply representing the component of the derivative of the displacement field in the radial direction.

B2 Term involving observer's velocity

Elaborating on the remaining part of equation (B7), in combination with equation (B8), yields the term involving the observer's velocity:

$$\Psi_s^{\text{obs}} \equiv \frac{\mathbf{v}_{\text{obs}}}{Ha} \cdot \hat{\mathbf{x}} \hat{\mathbf{x}}. \quad (\text{B17})$$

This term is valid for any structure formation model encoded in a displacement field Ψ .

$$(Ha) \frac{\partial \Psi_{s_k}^{\text{obs}}}{\partial \check{q}_i} = \mathbf{v}_k^{\text{obs}} \cdot \frac{\partial \hat{\mathbf{x}}_k}{\partial \check{q}_i} \hat{\mathbf{x}}_k + \mathbf{v}_k^{\text{obs}} \cdot \hat{\mathbf{x}}_k \frac{\partial \hat{\mathbf{x}}_k}{\partial \check{q}_i} \\ = \frac{1}{\|\mathbf{x}_k\|} \left[\mathbf{v}_k^{\text{obs}} \cdot \left(\frac{\partial \Psi_k}{\partial \check{q}_i} - \left(\hat{\mathbf{x}}_k \cdot \frac{\partial \Psi_k}{\partial \check{q}_i} \right) \hat{\mathbf{x}}_k \right) \right] \hat{\mathbf{x}}_k \\ + \frac{1}{\|\mathbf{x}_k\|} (\mathbf{v}_k^{\text{obs}} \cdot \hat{\mathbf{x}}_k) \left[\frac{\partial \Psi_k}{\partial \check{q}_i} - \left(\hat{\mathbf{x}}_k \cdot \frac{\partial \Psi_k}{\partial \check{q}_i} \right) \hat{\mathbf{x}}_k \right] \\ = \frac{1}{\|\mathbf{x}_k\|} \left[\mathbf{v}_k^{\text{obs}} \cdot \frac{\partial \Psi_k}{\partial \check{q}_i} \hat{\mathbf{x}}_k - 2(\mathbf{v}_k^{\text{obs}} \cdot \hat{\mathbf{x}}_k) \left(\frac{\partial \Psi_k}{\partial \check{q}_i} \cdot \hat{\mathbf{x}}_k \right) \hat{\mathbf{x}}_k + (\mathbf{v}_k^{\text{obs}} \cdot \hat{\mathbf{x}}_k) \frac{\partial \Psi_k}{\partial \check{q}_i} \right]. \quad (\text{B18})$$

This expression has the same structure as equation (B12), except for the first 'A' term. The latter disappears because it is constant, yielding

$$\frac{\partial \Psi_{s_k}^{\text{obs}}}{\partial \check{q}_i} \equiv \frac{1}{Ha} \left[(B_{ki}^{\text{obs}} - 2C_k^{\text{obs}} A_{ki}^{\text{obs}}) \hat{\mathbf{x}}_k + C_k^{\text{obs}} \frac{\partial \Psi_k}{\partial \check{q}_i} \right], \quad (\text{B19})$$

where we define

$$A_{ki}^{\text{obs}} \equiv \frac{\partial \Psi_k}{\partial \tilde{q}_i} \cdot \hat{\mathbf{x}}_k, \quad (\text{B20})$$

$$B_{ki}^{\text{obs}} \equiv \frac{1}{\|\mathbf{x}_k\|} \mathbf{v}_k^{\text{obs}} \cdot \frac{\partial \Psi_k}{\partial \tilde{q}_i}, \quad (\text{B21})$$

$$C_k^{\text{obs}} \equiv \frac{1}{\|\mathbf{x}_k\|} \mathbf{v}_k^{\text{obs}} \cdot \hat{\mathbf{x}}_k. \quad (\text{B22})$$

The similarity between the expressions reveals that it has the same behaviour of non-radial contributions near the observer as the Zel'dovich term.

B3 Plane parallel approximation

The above two terms can be greatly simplified by taking a plane parallel approximation of RSS, effectively putting the observer at an infinite distance. The $\frac{1}{\|\mathbf{x}\|}$ will then make the B and C terms go to zero, as well as the full observer's velocity term of the derivative of Ψ_s in equation (B19).

For the Zel'dovich model in plane parallel RSS, we are left with

$$\left. \frac{\partial \Psi_{sk}}{\partial \tilde{q}_i} \right|_{\text{plane parallel}} = \frac{\partial \Psi_{sk}^{(1)}}{\partial \tilde{q}_i} = f^{(1)} \left(\frac{\partial \Psi_k}{\partial \tilde{q}_i} \cdot \hat{\mathbf{x}}_k \right) \hat{\mathbf{x}}_k. \quad (\text{B23})$$

It is good to point out that Ψ_s and $\Psi_s^{(1)}$ are intrinsically different. Only in this specific case are their derivatives equal, as the observer's velocity term part is equal to zero.

Without loss of generality, we will restrict our studies to the plane parallel approximation for simplicity.

Finally, we insert this back into the $F^{\mathcal{L}, \text{rss}}$ term of equation (B6):

$$\begin{aligned} F_i^{\mathcal{L}, \text{rss}} &= -f^{(1)} \sum_k (\mathbf{V}_k \cdot \hat{\mathbf{x}}_k) \left(\frac{\partial \Psi_k}{\partial \tilde{q}_i} \cdot \hat{\mathbf{x}}_k \right) \\ &= -f^{(1)} \sum_k (\mathbf{V}_k \cdot \hat{\mathbf{x}}_k) \frac{1}{N} \sum_l e^{i\mathbf{k}_l \cdot \mathbf{q}_k} \left(-\frac{i\mathbf{k}_l}{k_l^2} \right) \\ &\quad \cdot \hat{\mathbf{x}}_k \sum_m e^{-i\mathbf{k}_l \cdot \mathbf{q}_m} \frac{\partial \vartheta_m}{\partial \tilde{q}_i} \\ &= -\frac{f^{(1)}}{N} \sum_m \sum_l \sum_k e^{-i\mathbf{k}_l \cdot \mathbf{q}_m} e^{i\mathbf{k}_l \cdot \mathbf{q}_k} (\mathbf{V}_k \cdot \hat{\mathbf{x}}_k) \left(-\frac{i\mathbf{k}_l}{k_l^2} \right) \\ &\quad \cdot \hat{\mathbf{x}}_k \frac{\partial \vartheta_m}{\partial \tilde{q}_i} \\ &\equiv -\frac{f^{(1)}}{N} \sum_m \sum_l \sum_k e^{-i\mathbf{k}_l \cdot \mathbf{q}_m} e^{i\mathbf{k}_l \cdot \mathbf{q}_k} \left(-\frac{i\mathbf{k}_l}{k_l^2} \right) \\ &\quad \cdot \mathbf{V}_k \frac{\partial \vartheta_m}{\partial \tilde{q}_i}, \end{aligned} \quad (\text{B24})$$

where we define \mathbf{V}_i^r as

$$\mathbf{V}_i^r \equiv (\mathbf{V}_i \cdot \hat{\mathbf{x}}_i) \hat{\mathbf{x}}_i. \quad (\text{B25})$$

We are left with almost the same result as in equation (3.27) of Bos (2016), the only difference being the replacement of \mathbf{V} with $f^{(1)} \mathbf{V}^r$. We can then conclude that in order to reuse equation (A10), we need to merely replace $h(\mathbf{q}_m)$ by $h^{\text{rss}}(\mathbf{q}_m)$, defined as

$$h^{\text{rss}}(\mathbf{q}_m) \equiv \nabla^{-2} \nabla \cdot (\mathbf{V}_m + f^{(1)} \mathbf{V}_m^r). \quad (\text{B26})$$

This results in

$$F_i^{\mathcal{L}} = \sum_m h_m^{\text{rss}} \frac{\partial \vartheta_m}{\partial \tilde{q}_i}. \quad (\text{B27})$$

In the implementation of this algorithm, it is convenient to choose the direction of the observer parallel to one of the coordinate axes. In this way, $\hat{\mathbf{x}}$ only changes one component of \mathbf{V} . This modification can easily be implemented in a pre-existing HMC sampling code.

APPENDIX C: ILLUSTRATION OF THE REDSHIFT SPACE PROBLEM

In this appendix, we illustrate the problems that one encounters when RSDs are not modelled. In summary, RSDs introduce an apparent anisotropic clustering in the data, which must be dealt with to reconstruct the real-space (initial and final) density field. When one does not correct for these distortions, the resulting reconstructions will be similarly distorted.

Note that we explicitly differentiate between the Kaiser effect and the effects we illustrate in this Appendix. Both in the Kaiser effect and in the effects in our ‘uncorrected RSS’ runs, the main issue is deformation of structures along the line of sight. However, the term *Kaiser effect* pertains specifically to the generated structure in RSS. In our case, the Kaiser effect is only present in the mock observations that we put into BARCODE. It is not present in the resulting reconstructed realizations because these are either in Lagrangian space or in Eulerian (real) space, not in RSS! The imprint of the Kaiser effect in these runs is visible because the wrong assumption was made that the observations are in real space, not in RSS. The algorithm then tries to find a set of initial conditions that reproduce the anisotropies purely through gravitational evolution, without any RSS assumptions. To clearly distinguish this imprint of the Kaiser effect on the reconstructions from the Kaiser effect itself in the future, we propose to call this effect the *Kaiser effect echo*.

C1 RSD artefacts in reconstructions

To show the problem with RSDs in observations, we obtained a set of regular Zel'dovich model samples with BARCODE. As with real data, the mock observational input is in RSS, *but we do not account for this in the model*. These are the ‘obs_rsd’ type runs that we mentioned in Section 3.3.

In Fig. C1, we first look at a slice of the Lagrangian density field. By eye, the results are strikingly similar to those with the RSS model. Structures are again well reconstructed in the mean of the reconstructions (top right) and the amount of variation around the mean is at a similar level (bottom panels).

We next look at the match of the corresponding ensemble average Eulerian density field to the true field in Fig. C2. The differences between the true (top left) and mean-reconstructed (top right) fields again are subtle and hard to discern by eye on paper. The best way to see the difference is to alternate between the mean reconstruction image of this figure and that of the RSS-model results in Fig. 7, laid on top of each other in an animation.⁸ In this way, it becomes immediately apparent that a strong anisotropic imprint is imparted on the reconstructions: along the line-of-sight direction (horizontal in the image), the voids are stretched and the overdense structures are compressed. In fact, when one compares this figure to the RSS

⁸ An animated GIF showing this is provided at http://egpbos.nl/rsd-eul_obs-eul/.

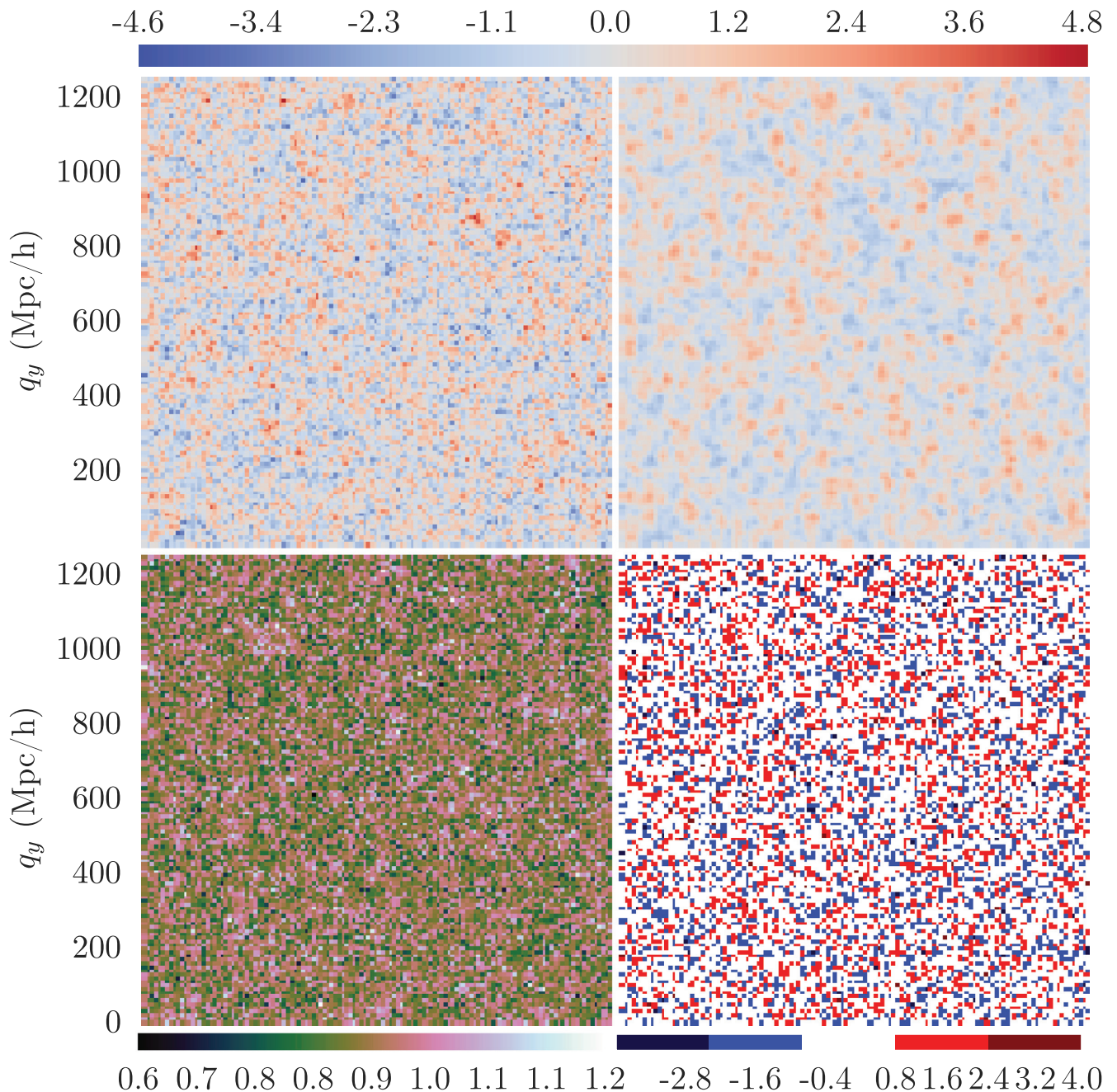


Figure C1. Mean sample *Lagrangian* densities (top right) compared to true (top left), together with std (bottom left), and the difference of the true and mean fields (bottom right).

density field of Fig. 2 the problem becomes apparent.⁹ When one does not model RSS, the algorithm converges on the true RSS field instead of the true Eulerian space field.

Our intention is to *remove* RSDs, not reproduce them. In the latter case, the corresponding Lagrangian density is quite incorrect as well. When one does not include RSDs and one uses data that is itself affected by RSDs, the model is simply incorrect. One assumes the RSS density distribution to equal the Eulerian space density. This begets a Lagrangian density that reproduces the observed RSDs as if they were true physical features of the large-scale

structure. Instead, they are optical illusions and must be treated accordingly.

The difference plot (bottom right) shows the shift of many features along the z -axis. A bi-polar pattern along this line-of-sight axis is apparent for most clusters.

C2 RSD artefacts in two-point statistics

The reconstruction problems are visible in the power spectrum as well, as can be seen in Fig. C3. This shows the boost in power at low k that is known as the Kaiser effect. In the difference plot (bottom panel of Fig. C3), this reveals itself as a downward turn. We want

⁹Animated GIF at http://egpbos.nl/rsd-rss_obs-eul/

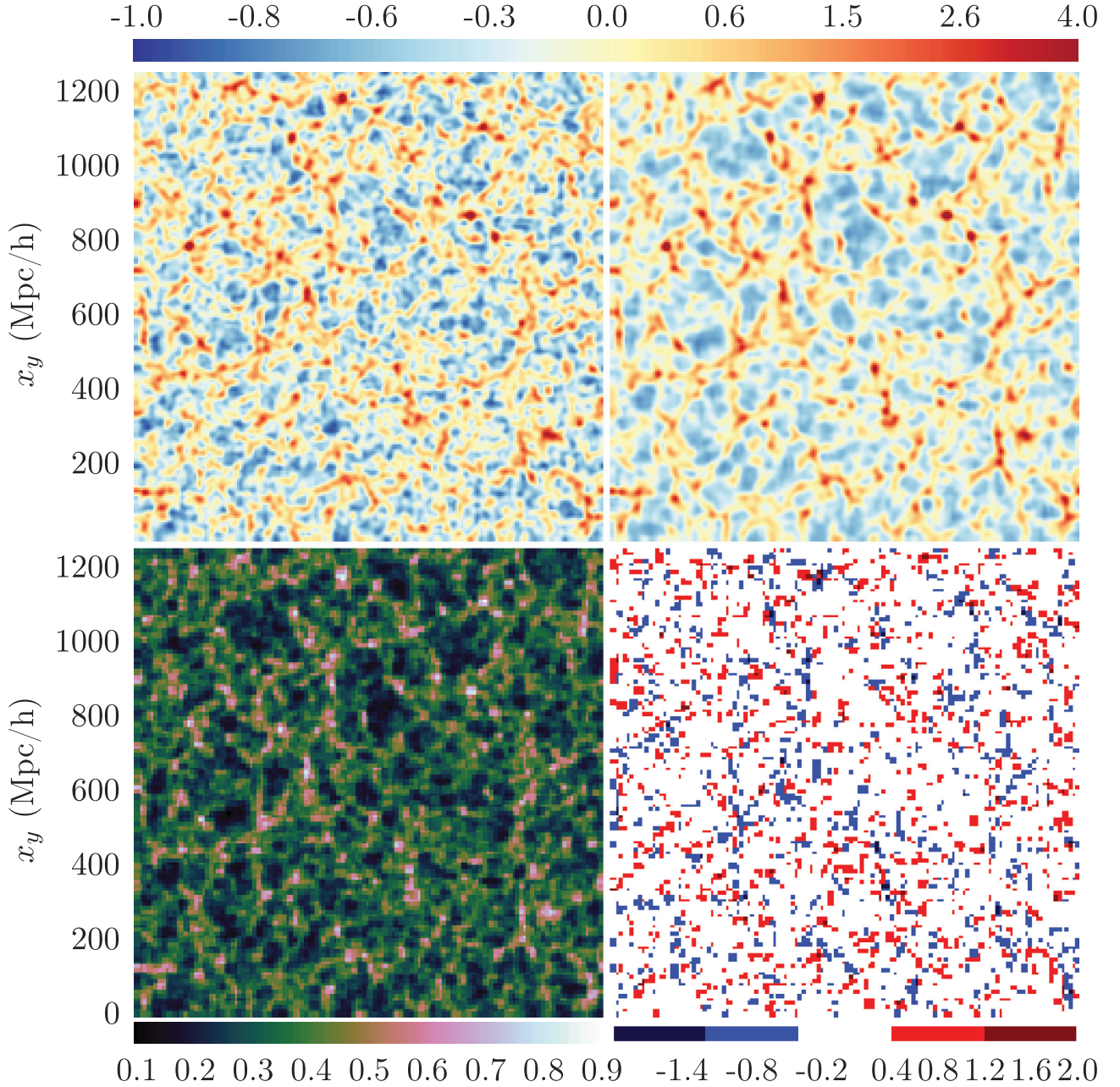


Figure C2. Mean sample *Eulerian* densities (top right) compared to true (top left), together with std (bottom left), and the difference of the true and mean fields (bottom right).

to get rid of this effect as well since we are trying to reconstruct RSD-less density fields.

In Fig. C4, we show 2D correlation functions (see Appendix E) that further illustrate the problem that is caused by having RSDs in the observations. What one would like to see is in the left-hand panel: the isotropic $\xi(r_{\perp}, r_{\parallel})$ of the true density field. The central panel shows the ensemble average of the samples that the algorithm

obtains when RSDs are present in the mock observations, but are not modelled.

The reconstructions show large anisotropy (non-circular $\xi(r_{\perp}, r_{\parallel})$). Note again that we only see here the effect of the coherent flow RSDs, which cause ‘great walls’. This is seen in the horizontal (perpendicular to line of sight) stretching of the otherwise circular correlation function.

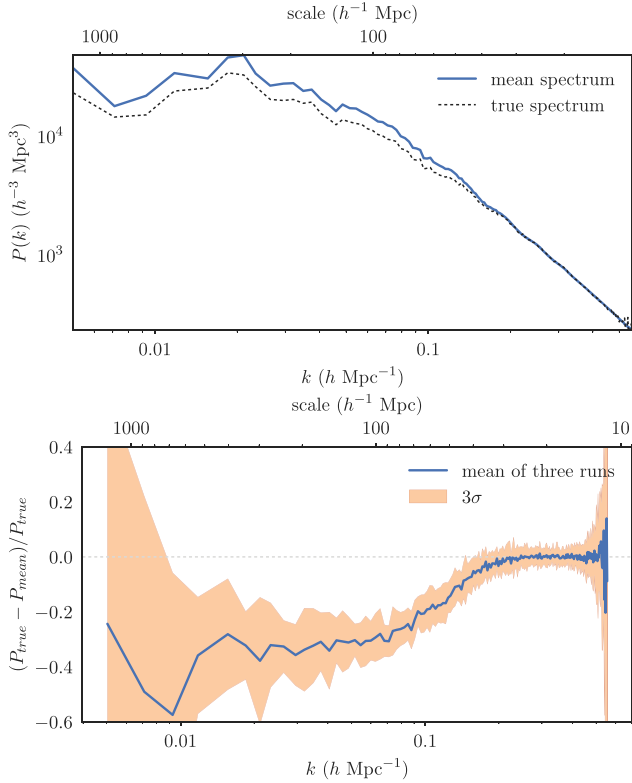


Figure C3. Sample Lagrangian power spectra compared to true; true compared to the mean and standard deviation of the sampled densities. Note that the difference plot (bottom panel) uses samples from three different runs with different true fields, while the power spectrum plot itself (top panel) is only from one true field.

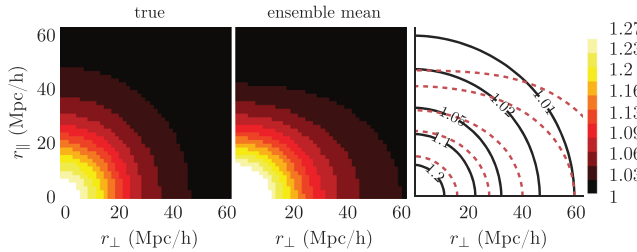


Figure C4. Eulerian space 2D correlation functions: true (left) versus ensemble mean (centre). The right-hand panel shows the two functions in contours: the solid black for the true function and the dashed red for the ensemble mean.

APPENDIX D: MCMC CHAIN PERFORMANCE

Some questions remain regarding performance of the algorithm given our new model, in terms of both speed and accuracy. Here, we formulate some answers to the questions of how the chain evolves compared to the case without RSDs presented in Bos (2016) and in Appendix C and whether we see any remaining imprints of the RSDs.

D1 Burn-in phase

When trying to analyse chain mixing speed, one obvious place to begin is the burn-in phase. In Fig. D1, we compare the likelihood of three different chains for each of the two models, the Zel'dovich

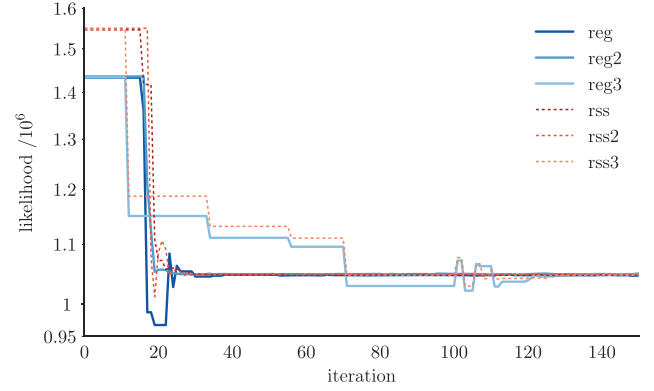


Figure D1. Likelihood during the burn-in phase of six chains. Results are shown for the ‘regular’ Eulerian space model in the solid blue lines and the ‘rss’ model in redshift space in the dashed red lines. For each model, the three lines correspond to runs with three different true Lagrangian initial fields, generated with three different random number generator seeds.

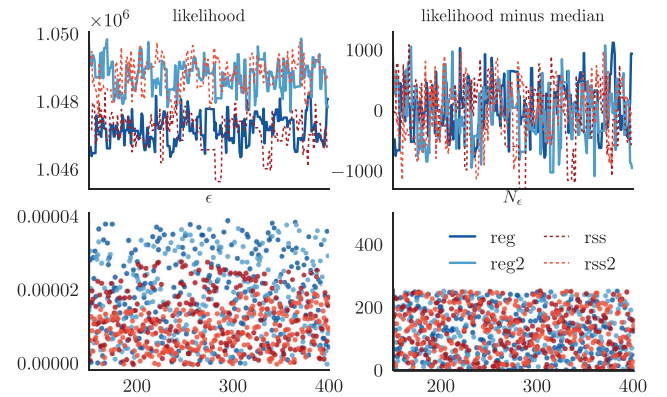


Figure D2. Chain parameter evolution versus iteration number (the horizontal axis). The iteration number includes both accepted and rejected steps. The top panels show the likelihood (the right-hand panel with the median of the likelihood subtracted for easier comparison), the bottom left-hand panel shows ϵ , and the bottom right shows N_ϵ . Regular runs in the solid blue shades; redshift model runs in the dashed red shades.

model in Eulerian and in RSS. From this, we conclude that burn-in does not seem at all slower. The runs seem equally fast in Eulerian and in RSS. That also means that defining step 150 as the cut-off for burn-in is acceptable for the RSS model as well.

D2 Chain performance after burn-in and adaptive ϵ evaluation

In Fig. D2, we show chain parameters for the whole chain (after burn-in). Note that the iteration number on the horizontal axis includes both accepted and rejected steps. The variation in the likelihood can be used as a measure of the variability of the chain, as it compares the sample to the input field. Two Eulerian space and two RSS chains are shown in these figures.

The variation in likelihood is nearly identical for the regular and RSS chains, with a standard deviation of slightly over 400 for both chains. This is more directly illustrated in the top right-hand panel, where we subtract the median value of the chain, so that the actual variation around the median can be compared for the different chains. The random patterns are statistically indistinguishable.

In the bottom panel, we additionally show ϵ (bottom left) and N_ϵ (bottom right). Here, finally the expected difference in the chain

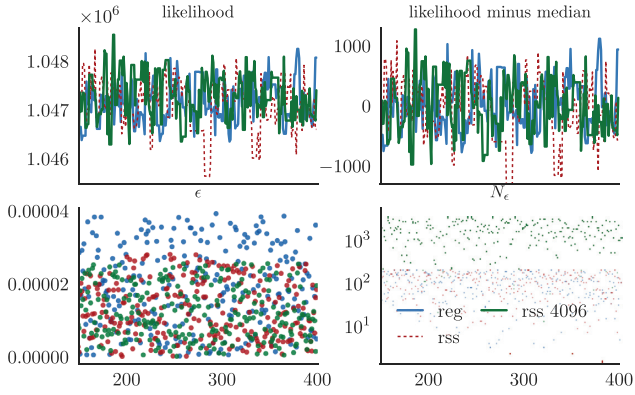


Figure D3. Chain parameter evolution versus iteration number (the horizontal axis) for different N_ϵ values. The iteration number includes both accepted and rejected steps. The top panels show the likelihood (the right-hand panel with the median of the likelihood subtracted for easier comparison), the bottom left-hand panel shows ϵ , and the bottom right shows N_ϵ . Regular run in solid blue, redshift model runs (including the two with a higher N_ϵ) in the dashed red shades.

performance between the regular and rss runs shows. The leap-frog step-size adapts to smaller values in the RSS model runs. This means that for some reason the orbit through the posterior's parameter space accrues more errors, i.e. its Hamiltonian is more sensitive to the exact path taken. The RSS model constrains the sampler less tightly since it allows for more possible configurations along the line of sight than when using a regular space model. This means that it is easier for the sampler to end up in a location that fits less well with the true field than in the regular model case. This explains why a shorter leap-frog path may be necessary in the RSS model.

D3 Optimizing number of leap-frog steps

In Bos (2016), we found that for runs with $L = 200 h^{-1}\text{Mpc}$ and $N_x = 64$, a higher number of leap-frog steps N_ϵ than 256 improved the performance of the chain in that its samples covered a larger part of the posterior's parameter space. We tried a higher value in this work as well to test whether we have reached optimal chain performance.

In Fig. D3, we show the results of another RSS model runs with a higher N_ϵ value of 4096. This significantly higher value does not seem to make any noticeable difference in the chain evolution parameters. It does make the code slower, as it scales with N_ϵ . We can conclude that 256 leap-frog steps per iteration are plenty for proper chain performance.

APPENDIX E: 2D CORRELATION FUNCTIONS

The stronger clustering implied by the Kaiser effect results in a boost of the contrast of walls and filaments perpendicular to the line of sight. This boost is reflected in the higher amplitude of the power spectrum. However, it does not differentiate between the line of sight and the other two, unaffected directions. A common statistic that does take this into account is the 2D (sky – redshift) correlation function. We give a short recap on this statistic.

In general, the correlation function $\xi(\mathbf{x}_A, \mathbf{x}_B)$ of point A with point B in a homogeneous field is dependent only on the distance vector \mathbf{r}_{AB} between the two points, i.e.

$$\xi(\mathbf{x}_A, \mathbf{x}_B) = \xi(\mathbf{x}_A - \mathbf{x}_B) \equiv \xi(\mathbf{r}_{AB}). \quad (\text{E1})$$

Often, one also assumes the field to be isotropic, which further reduces the dependence of the correlation function to only the (absolute) distance scalar between two points:

$$\xi(\mathbf{r}_{AB}) = \xi(|\mathbf{r}_{AB}|) \equiv \xi(r_{AB}). \quad (\text{E2})$$

The assumptions of homogeneity and isotropy form the basis of the FLWR¹⁰ metric of space-time.

When dealing with correlation functions of fields with RSDs, the latter assumption no longer holds. The RSDs create a marked anisotropy between the line of sight and the sky. The isotropy in the plane of the sky remains preserved. This means that for correlation functions of density fields affected by RSDs, we can reduce the correlation function to a function of two scalar variables instead of one: one in the direction of the line of sight, and one in the remaining two (transverse) directions. In our simplified case, where the observer is far away, the line-of-sight direction can be (and is) chosen to be along the z -axis. This gives us

$$\xi(\mathbf{r}) = \xi(|\mathbf{r} - r_z \hat{z}|, r_z) \equiv \xi(r_\perp, r_\parallel), \quad (\text{E3})$$

where $r_z = \mathbf{r} \cdot \hat{z}$ and for simplicity we drop the AB subscripts.

In the literature, often r_\parallel is called the radial pair separation π and r_\perp is called the transverse pair separation σ (Peacock et al. 2001). Note that the parallel (to the line-of-sight) axis, which in our case is the z -axis, is vertical in the $\xi(r_\perp, r_\parallel)$ plots, but horizontal in the density slice plots. The former was chosen to correspond to the way $\xi(r_\perp, r_\parallel)$ plots are usually shown in the literature (see e.g. Hawkins et al. 2003).

The Kaiser effect is reflected in the 2D correlation functions. In this case, only in the transverse direction. The increased correlation on large scales corresponds to the ‘great walls’ that are enhanced in the observations (Sargent & Turner 1977; Ryden & Melott 1996; Thomas et al. 2004). While the difference appears to be in the transverse direction, the cause of this flattening is the compression along the line-of-sight direction. Due to this compression, both a modest amount of flattening in the radial direction and the more apparent widening into the transverse direction are brought about. The compression also produces a generally higher level of correlation.

Each quadrant contains exactly the same information since $\xi(r_\perp, r_\parallel)$ is symmetric along its two axes. In this paper, we plot the upper left quadrant only, which offers greater detail at the expense of making it slightly harder to see the shape.

APPENDIX F: KERNEL DENSITY ESTIMATION

The derivative of the density field $\frac{\partial \rho(\mathbf{x}_i)}{\partial \mathbf{x}(\mathbf{q})}$ pops up in the derivative of the log-likelihood of equation (B1). This must be treated very carefully because a mixing of coordinate systems takes place here. We take the derivative of the Eulerian density field on a regular grid defined by the observations, but we derive to an irregular $\mathbf{x}(\mathbf{q})$ coordinate. Even though \mathbf{q} is on a regular (initial Lagrangian density field) grid, $\mathbf{x}(\mathbf{q})$ is certainly not. This means we cannot simply put the density on a grid and then take a derivative (easily done numerically in Fourier space by multiplying by $i\mathbf{k}$). One might think that by interpolating the regularized field $\rho(\mathbf{x}_i)$ on locations $\mathbf{x}(\mathbf{q})$ one could also overcome this, but this is not the case. In fact, this introduces severe instabilities in the algorithm.

As always for HMC algorithms, the only solution is to keep every calculation as explicit as possible. We do this by formulating the

¹⁰Friedmann, Lemaître, Walker, and Robertson.

density field in terms of particles with a kernel that distributes their mass over space. This allows us to analytically express the derivative to the particle kernel necessary in the likelihood force. The following general expression for density $\rho(\mathbf{x})$ then applies:

$$\rho(\mathbf{x}) = \sum_{\text{particles } i} m_i W(\mathbf{x} - \mathbf{x}_i), \quad (\text{F1})$$

given a kernel $W(\mathbf{x}_i, \mathbf{x})$ that gives the density contribution of the particle at \mathbf{x} given a particle i at \mathbf{x}_i , and a particle mass m_i to the continuous density field ρ at \mathbf{x} .¹¹ Our mass will be defined by the critical density ρ_c , box volume V , and number of grid cells N , as we are only concerned with equal mass Lagrangian particles:

$$m_i = \rho_c \frac{V}{N}. \quad (\text{F2})$$

This kernel formulation conveniently integrates the particle nature of the transformation from Lagrangian to Eulerian space; this is, after all, a hydrodynamical representation of matter fields, based on particles.

One could in principle convolve fields with these kernels to smooth them out, or deconvolve fields with them by dividing by them in Fourier space. Wang et al. (2013) suggest that the latter is necessary when using kernel-based density estimation and/or interpolation methods. We do not follow them in this, as we discuss in Bos (2016).

For this work, we chose to use an SPH spline kernel (Monaghan 1992; Springel 2010) for each particle at $\mathbf{x}(\mathbf{q})$. The advantages of this kernel are

- (i) smoothness (it is a second-order Taylor expansion of a Gaussian), which also gives us nicely behaving derivatives (first and second), and
- (ii) boundedness within a few neighbouring cells, which makes it far more computationally efficient than, e.g. a Gaussian that has infinite extent.

Other options, such as Gaussian, NGP, CIC, and TSC kernels only have one of the above properties, not both. We can also not use the more precise DTFE scheme (Schaap & van de Weygaert 2000; Cautun & van de Weygaert 2011; van de Weygaert & Platen 2011), because it is not kernel-based, making analytical derivatives a lot harder to calculate.

The SPH spline kernel is defined in 3D following Monaghan (1992):

$$W(q; h_s) = \frac{1}{\pi h_s^3} \begin{cases} 1 - \frac{3}{2}q^2 + \frac{3}{4}q^3 & \text{if } 0 \leq q < 1 \\ \frac{1}{4}(2 - q)^3 & \text{if } 1 \leq q < 2 \\ 0 & \text{otherwise,} \end{cases} \quad (\text{F3})$$

where h_s is the scale parameter of the kernel that defines its size and q is given by

$$q \equiv \frac{|\mathbf{x} - \mathbf{x}_i|}{h_s}. \quad (\text{F4})$$

¹¹To transform from the \mathbf{q} to the \mathbf{x} grid, we start by evolving particles forward in time to the desired redshift using the structure formation model we chose (Zel'dovich in this work), given the input field $\delta(\mathbf{q})$. The resulting particle distribution in \mathbf{x} comoving coordinate space is then converted to a new regularly gridded density field $\rho(\mathbf{x})$ by means of an SPH field estimation scheme.

An important property of the kernel thus normalized is that, independent of the choice of h_s , we have

$$\int_V W(q; h_s) d\mathbf{x} = 1, \quad (\text{F5})$$

which is necessary for equation (F1) to be correct. Note that the integral is unitless, as W 's units are Mpc^{-3} ; W is effectively a number density that spreads the mass of the Lagrangian tracer particle i at \mathbf{x}_i over its surroundings.

We need the gradient of this kernel with respect to $\mathbf{x}_i = \mathbf{x}(\mathbf{q}_i)$ to calculate the likelihood force term. This gradient is given by

$$\frac{\partial W(\mathbf{x} - \mathbf{x}_i; h_s)}{\partial \mathbf{x}_i} = \frac{1}{\pi h_s^5} \frac{\mathbf{x}_i - \mathbf{x}}{q} \begin{cases} \frac{9}{4}q^2 - 3q & \text{if } 0 \leq q \leq 1 \\ -\frac{3}{4}(2 - q)^2 & \text{if } 1 \leq q \leq 2 \\ 0 & \text{otherwise} \end{cases}, \quad (\text{F6})$$

where in the $F^{\mathcal{L}}$ equations \mathbf{x}_i is replaced with $\mathbf{x}(\mathbf{q}_i)$ and \mathbf{x} with \mathbf{x}_j , which are the terms we will actually use in the formulas (though it works for any combination of \mathbf{x} positions).

APPENDIX G: SECOND-ORDER LAGRANGIAN PERTURBATION THEORY EXTENSION

In the implementation used in this paper, BARCODE uses the Zel'dovich or first-order LPT model as its structure formation model. This model has the advantage of giving a decent match to reality over a range of scales; from large to semilinear scales. On large scales, however, the 2LPT model of structure formation is more accurate, reproducing higher order structural statistics better than, for instance, the ZA (Neyrinck 2013). Also, it takes into account the second-order tidal terms, which is important for overall large-scale structure formation. In this appendix (Appendix G2), we derive the equations necessary to include the 2LPT model in BARCODE.

While 2LPT works well at large scales, it fails badly at smaller scales. Its main downside is that it produces even stronger overshoot in clustered regions than the Zel'dovich model. Fig. 6 from Neyrinck (2013) illustrates this fact succinctly.

To correct 2LPT on small scales, we adopt the *Augmented LPT* model of Kitaura & Heß (2013). This model adds an SC component to the 2LPT model. Through this combination, the massive overshoot is removed completely. Small scales are corrected for in a way that is still analytically tractable and thus usable in our Hamiltonian sampler. We will refer to this model by the more explicit acronym '2LPT + SC' instead of the previously used ALPT. In Appendix G3, we derive the equations necessary to implement 2LPT + SC in BARCODE.

This model, though a great improvement to bare 2LPT, is not without its shortcomings. One of particular interest is the fact that it is, by nature, spherical. However, structure formation as a process is driven by anisotropic collapse (Zel'dovich 1970; Icke 1973). Since this is what we study, we need to assess the impact of the addition of an SC model to anisotropic measures of large-scale structure. The validity of these models with respect to important aspects of the large-scale structure will be tested in a future study.

G1 Notation

Most of this appendix uses the same definitions and the same notation as the rest of this paper. However, we must make some additional remarks on index notation. Normal indices usually start

from i in the alphabet and indicate grid cells. We use these indices when talking about Fourier transforms, for instance. Underlined indices start from a , i.e. \underline{a} . These are used when talking about components of regular vectors such as coordinate vectors \mathbf{x} , velocities \mathbf{v} , or a displacement vector Ψ . Some quantities have a combination of these two types of indices, e.g. when talking about (Fourier transforms of) discretized vector fields, such as the i th cell of the \underline{a} th component of a displacement field Ψ : $\Psi_{i\underline{a}}$.

When we use matrices, two indices are necessary. The same goes for the Kronecker delta δ_{ij}^K . When, instead of indices, the full coordinates are written in the subscript of the Kronecker delta, e.g. $\delta_{\mathbf{q}_i, \mathbf{q}_j}^K$, a comma is put in between for clarity. We clarify this, since, in other cases, a comma in the subscript means a derivative with respect to the coordinates corresponding to the indices (for instance, in equation G4).

G2 Second-order Lagrangian perturbation theory

2LPT adds a second-order term to the equation for ϑ (Bernardeau et al. 2002, equation 94):

$$\vartheta^{(2)}(\mathbf{q}) = D_2 \Delta^{(2)}, \quad (\text{G1})$$

so that the full equation for ϑ in 2LPT reads

$$\vartheta = -D_1 \Delta^{(1)} + D_2 \Delta^{(2)}. \quad (\text{G2})$$

The second-order spatial term is dependent on the first-order term from equation (A19). This allows us to take the derivative to $\delta(\mathbf{q}_i) = \delta^{(1)}(\mathbf{q}_i)$, obtaining

$$\frac{\partial \vartheta(\mathbf{q})}{\partial \delta(\mathbf{q}_i)} = -\delta_{\mathbf{q}, \mathbf{q}_i}^K + D_2 \frac{\partial \Delta^{(2)}}{\partial \delta(\mathbf{q}_i)}. \quad (\text{G3})$$

The second-order term of the density in LPT, $\Delta^{(2)}$, is given by

$$\Delta^{(2)} \equiv \sum_{\underline{a} > \underline{b}} \left(\phi_{,\underline{a}\underline{a}}^{(1)}(\mathbf{q}) \phi_{,\underline{b}\underline{b}}^{(1)}(\mathbf{q}) - \left(\phi_{,\underline{a}\underline{b}}^{(1)}(\mathbf{q}) \right)^2 \right), \quad (\text{G4})$$

where the comma- $\underline{a}/\underline{b}$ -subscripts denote the derivative to the \underline{a} th/ \underline{b} th component of the coordinates (x, y, z). $\phi^{(1)}$ is the first-order term in the LPT gravitational potential expansion, which is found by solving the Poisson equation for the first-order terms:

$$\nabla_{\mathbf{q}}^2 \phi^{(1)}(\mathbf{q}) = \Delta^{(1)}. \quad (\text{G5})$$

For the derivative of $\Delta^{(2)}$ in equation (G3), one then obtains

$$\begin{aligned} \frac{\partial \Delta^{(2)}}{\partial \delta(\mathbf{q}_i)} &= \sum_{\underline{a} > \underline{b}} \left(\phi_{,\underline{a}\underline{a}}^{(1)}(\mathbf{q}) \frac{\partial \phi_{,\underline{b}\underline{b}}^{(1)}(\mathbf{q})}{\partial \delta(\mathbf{q}_i)} + \phi_{,\underline{b}\underline{b}}^{(1)}(\mathbf{q}) \frac{\partial \phi_{,\underline{a}\underline{a}}^{(1)}(\mathbf{q})}{\partial \delta(\mathbf{q}_i)} \right. \\ &\quad \left. - 2 \phi_{,\underline{a}\underline{b}}^{(1)}(\mathbf{q}) \frac{\partial \phi_{,\underline{a}\underline{b}}^{(1)}(\mathbf{q})}{\partial \delta(\mathbf{q}_i)} \right). \end{aligned} \quad (\text{G6})$$

The derivative of $\phi_{,\underline{a}\underline{b}}^{(1)}$ to $\delta(\mathbf{q}_i) = D_1 \Delta^{(1)}(\mathbf{q}_i)$ can be solved in Fourier-space. The Poisson equation in Fourier-space can be written as

$$\hat{\phi}^{(1)}(\mathbf{k}) = -\frac{1}{k^2} \hat{\Delta}^{(1)}(\mathbf{k}), \quad (\text{G7})$$

which gives us this relation for $\phi^{(1)}$:

$$\phi^{(1)}(\mathbf{q}) = -\frac{1}{N} \sum_j e^{-i\mathbf{k}_j \cdot \mathbf{q}} \frac{1}{k_j^2} \hat{\Delta}^{(1)}(\mathbf{k}_j). \quad (\text{G8})$$

Taking the derivatives to the \underline{a} th and \underline{b} th coordinate components,

$$\begin{aligned} \phi_{,\underline{a}\underline{b}}^{(1)}(\mathbf{q}) &= \partial_{\underline{a}} \partial_{\underline{b}} \phi^{(1)}(\mathbf{q}) \\ &= \frac{1}{N} \sum_j e^{i\mathbf{k}_j \cdot \mathbf{q}} \frac{k_{ja} k_{jb}}{k_j^2} \left[\sum_{\mathbf{q}'} e^{-i\mathbf{k}_j \cdot \mathbf{q}'} \Delta^{(1)}(\mathbf{q}') \right], \end{aligned} \quad (\text{G9})$$

and the derivative of this to $\delta(\mathbf{q}_i)$, we end up with

$$\begin{aligned} \frac{\partial \phi_{,\underline{a}\underline{b}}^{(1)}(\mathbf{q})}{\partial \delta(\mathbf{q}_i)} &= \frac{1}{N} \sum_j e^{i\mathbf{k}_j \cdot \mathbf{q}} \frac{k_{ja} k_{jb}}{k_j^2} \left[\sum_{\mathbf{q}'} e^{-i\mathbf{k}_j \cdot \mathbf{q}'} \frac{\partial \Delta^{(1)}(\mathbf{q}')}{\partial D_1 \Delta^{(1)}(\mathbf{q}_i)} \right] \\ &= \frac{1}{D_1 N} \sum_j e^{i\mathbf{k}_j \cdot \mathbf{q}} \frac{k_{ja} k_{jb}}{k_j^2} \left[\sum_{\mathbf{q}'} e^{-i\mathbf{k}_j \cdot \mathbf{q}'} \delta_{\mathbf{q}', \mathbf{q}_i}^K \right] \\ &= \frac{1}{D_1 N} \sum_j e^{i\mathbf{k}_j \cdot \mathbf{q}} \frac{k_{ja} k_{jb}}{k_j^2} e^{-i\mathbf{k}_j \cdot \mathbf{q}_i}. \end{aligned} \quad (\text{G10})$$

The first exponential we can use, in the end, to get rid of the sum in equation (A9), by turning them into a DFT. The second exponential will be used with the sum over j in this equation. Note that for this equation to be non-zero, the \mathbf{q}_i and \mathbf{q} grids need to be the same! If they are not, the Kronecker delta will only equal zero. This final step thus necessarily couples the \mathbf{q}_i -grid of the signal, the $\delta(\mathbf{q})$ field sample, and that of the $\mathbf{x}(\mathbf{q})$ grid used in the likelihood \mathcal{L} .

Now, to simplify equation (G6), we can write each of the terms in the $\underline{a} > \underline{b}$ sum in a general way, writing them as specific instances of a general four-index quantity. Including the multiplication by $h(\mathbf{q}_m)$ and the sum over index m from equation (A9), and using the result from equation (G10), we then get for each generalized part (note the indices $\underline{a}, \underline{b}, \underline{c}$, and \underline{d}) of the full term:

$$\begin{aligned} \sum_m h(\mathbf{q}_m) D_2 \phi_{,\underline{c}\underline{d}}^{(1)}(\mathbf{q}_m) \frac{\partial \phi_{,\underline{a}\underline{b}}^{(1)}(\mathbf{q}_m)}{\partial \delta(\mathbf{q}_i)} \\ &= \sum_m h(\mathbf{q}_m) \frac{D_2}{D_1} \frac{1}{N} \phi_{,\underline{c}\underline{d}}^{(1)}(\mathbf{q}_m) \sum_j e^{i\mathbf{k}_j \cdot \mathbf{q}_m} \frac{k_{ja} k_{jb}}{k_j^2} e^{-i\mathbf{k}_j \cdot \mathbf{q}_i} \\ &\equiv \frac{D_2}{D_1} \frac{1}{N} \sum_j \hat{\alpha}_{\underline{c}\underline{d}}^*(\mathbf{k}_j) \frac{k_{ja} k_{jb}}{k_j^2} e^{-i\mathbf{k}_j \cdot \mathbf{q}_i} \\ &\equiv \frac{D_2}{D_1} \xi_{\underline{c}\underline{d}}^{*ab}(\mathbf{q}_i) \\ &= \frac{D_2}{D_1} \xi_{\underline{c}\underline{d}}^{ab}(\mathbf{q}_i), \end{aligned} \quad (\text{G11})$$

where we define the convenience functions $\alpha(\mathbf{q})$ and $\xi(\mathbf{q})$ as

$$\alpha_{\underline{c}\underline{d}}(\mathbf{q}) \equiv \left(h(\mathbf{q}) \phi_{,\underline{c}\underline{d}}^{(1)}(\mathbf{q}) \right)^* = h(\mathbf{q}) \phi_{,\underline{c}\underline{d}}^{(1)}(\mathbf{q}), \quad (\text{G12})$$

$$\begin{aligned} \hat{\xi}_{\underline{c}\underline{d}}^{ab}(\mathbf{k}) &\equiv \left(\frac{k_a k_b}{k^2} \hat{\alpha}_{\underline{c}\underline{d}}^*(\mathbf{k}) \right)^* = \frac{k_a k_b}{k^2} \hat{\alpha}_{\underline{c}\underline{d}}(\mathbf{k}) \\ \Rightarrow \xi_{\underline{c}\underline{d}}^{ab}(\mathbf{q}) &= \partial_{\underline{a}} \partial_{\underline{b}} \nabla^{-2} \left(h(\mathbf{q}) \phi_{,\underline{c}\underline{d}}^{(1)}(\mathbf{q}) \right), \end{aligned} \quad (\text{G13})$$

where we get rid of the conjugations because $\phi_{,\underline{c}\underline{d}}^{(1)}$ is real, as it is a physical term, $h(\mathbf{q})$ is also real, as previously noted, and the \mathbf{k} 's are real scale vectors. The double conjugation on $\hat{\alpha}_{\underline{c}\underline{d}}$ cancels out as well.

We will see $\xi_{\underline{c}\underline{d}}^{ab}$ again in Appendix G3, but in a slightly different form. This prompts us to use a functional instead, which we can use

again later on and which simplifies our code by only writing the general form of the function once:

$$\chi_{\underline{cd}}^{ab}(f(\mathbf{q}); \mathbf{q}) \equiv \partial_{\underline{a}} \partial_{\underline{b}} \nabla^{-2} \left(f(\mathbf{q}) \phi_{\underline{cd}}^{(1)}(\mathbf{q}) \right). \quad (\text{G14})$$

Then, finally, we can write down a worked out 2LPT version of the likelihood force (equation A9), combining the first- and second-order results:

$$F_i^{\mathcal{L}, 2\text{LPT}} = -h(\mathbf{q}_i) + \frac{D_2}{D_1} \sum_{\underline{a} > \underline{b}} \left(\chi_{\underline{bb}}^{aa} + \chi_{\underline{aa}}^{bb} - 2\chi_{\underline{ab}}^{ab} \right) \times (h(\mathbf{q}_i); \mathbf{q}_i). \quad (\text{G15})$$

The argument $(h(\mathbf{q}_i); \mathbf{q}_i)$ applies to all three χ s.

G3 2LPT and spherical collapse

For the 2LPT + SC model, the divergence of the displacement field $\vartheta(\delta^{(1)}(\mathbf{q}))$ is approximated by the weighted sum of a 2LPT part on large scales and an SC part on small scales, where the scales are split using Gaussian convolution (meaning that there is not an absolute splitting of scales, there will be overlap):

$$\begin{aligned} \vartheta &\equiv [K_G * (\nabla \cdot \Psi^{2\text{LPT}})] + (\nabla \cdot \Psi^{\text{SC}}) - [K_G * (\nabla \cdot \Psi^{\text{SC}})] \\ &= [K_G * (-D_1 \Delta^{(1)} + D_2 \Delta^{(2)})] + 3 \left(\sqrt{1 - \frac{2}{3} \delta} - 1 \right) \\ &\quad - \left[K_G * \left(3 \left(\sqrt{1 - \frac{2}{3} \delta} - 1 \right) \right) \right], \end{aligned} \quad (\text{G16})$$

where $K_G(\mathbf{q})$ is a Gaussian kernel at some characteristic scale R_G . The convolution (denoted by $*$) with the 2LPT part preserves mainly the scales above R_G and the other with ‘ $1 - K_G$ ’ preserves scales below that for the SC part. All terms are defined on the Lagrangian \mathbf{q} grid (we left out the function arguments for brevity).

The derivative we need can be identified as

$$\begin{aligned} \frac{\partial \vartheta(\mathbf{q})}{\partial \delta(\mathbf{q}_i)} &= \left[K_G * \left(-K_i^K + D_2 \frac{\partial \Delta^{(2)}}{\partial \delta(\mathbf{q}_i)} \right) \right](\mathbf{q}) \\ &\quad + \left(1 - \frac{2}{3} \delta(\mathbf{q}) \right)^{-\frac{1}{2}} K_i^K(\mathbf{q}) \\ &\quad - \left[K_G * \left(\left(1 - \frac{2}{3} \delta(\mathbf{q}) \right)^{-\frac{1}{2}} K_i^K \right) \right](\mathbf{q}). \end{aligned} \quad (\text{G17})$$

Here, we use an unusual notation for Kronecker delta:

$$K_i^K(\mathbf{q}) \equiv \delta_{\mathbf{q}, \mathbf{q}_i}^K. \quad (\text{G18})$$

This helps us to neatly write the Kronecker delta in convolution notation. The 2LPT part of this equation has been worked out in Appendix G2, so we recycle those results next. The convolutions with the Gaussian are also most efficiently done in (discrete) Fourier space. Note that the conversion between continuous and discrete FTs causes the following relation to contain a small error (a convolution of a Gaussian with a general function $f(\mathbf{q})$):

$$\begin{aligned} [K_G * f](\mathbf{q}) &= \frac{1}{N} \sum_j e^{ik_j \cdot \mathbf{q}} \hat{K}_G(\mathbf{k}_j) \hat{f}(\mathbf{k}_j) \\ &= \frac{1}{N} \sum_j e^{ik_j \cdot \mathbf{q}} e^{-\frac{k_j^2 R_G^2}{2}} \hat{f}(\mathbf{k}_j). \end{aligned} \quad (\text{G19})$$

In what follows, we will, term by term, reduce equation (A9), combined with the 2LPT + SC prescription of this section, to a series of Fourier transforms of several intermediate convenience functions. The numbered terms that form

$$F_i^{\mathcal{L}, 2\text{LPT} + \text{SC}} = \textcircled{1} + \textcircled{2} + \textcircled{3} + \textcircled{4} \quad (\text{G20})$$

are

$$\begin{aligned} \textcircled{1} &= \sum_m h(\mathbf{q}_m) [K_G * (-\delta_{\mathbf{q}_m, \mathbf{q}_i}^K)], \\ \textcircled{2} &= \sum_m h(\mathbf{q}_m) \left[K_G * \left(D_2 \frac{\partial \Delta^{(2)}(\mathbf{q}_m)}{\partial \delta(\mathbf{q}_i)} \right) \right] \equiv \sum_m h(\mathbf{q}_m) [\textcircled{*}], \\ \textcircled{3} &= \sum_m h(\mathbf{q}_m) \left(1 - \frac{2}{3} \delta(\mathbf{q}_m) \right)^{-\frac{1}{2}} \delta_{\mathbf{q}_m, \mathbf{q}_i}^K, \\ \textcircled{4} &= - \sum_m h(\mathbf{q}_m) \left[K_G * \left(\left(1 - \frac{2}{3} \delta(\mathbf{q}_m) \right)^{-\frac{1}{2}} \delta_{\mathbf{q}_m, \mathbf{q}_i}^K \right) \right]. \end{aligned}$$

G3.1 Term 1

This term contains, on first hand, a somewhat unintuitive convolution with a Kronecker delta function. After some manipulation, however, this term will quickly conform to something more in line with our comfortable view of the world.

Using the DFT of the Kronecker delta,

$$\mathcal{F}_D [\delta_{\mathbf{x}, \mathbf{x}'}^K] = \sum_{\mathbf{x}} e^{-ik \cdot \mathbf{x}} \delta_{\mathbf{x}, \mathbf{x}'}^K = e^{-ik \cdot \mathbf{x}'}, \quad (\text{G21})$$

we reshape term $\textcircled{1}$ into a simple convolution:

$$\begin{aligned} \textcircled{1} &= \sum_m h(\mathbf{q}_m) [K_G * (-K_i^K)](\mathbf{q}_m) \\ &= -\frac{1}{N} \sum_m h(\mathbf{q}_m) \sum_j e^{ik_j \cdot \mathbf{q}_m} \hat{K}_G(\mathbf{k}_j) e^{-ik_j \cdot \mathbf{q}_i} \\ &= -\frac{1}{N} \sum_j \sum_m h(\mathbf{q}_m) e^{ik_j \cdot \mathbf{q}_m} \hat{K}_G(\mathbf{k}_j) e^{-ik_j \cdot \mathbf{q}_i} \\ &\equiv -\frac{1}{N} \sum_j \hat{c}^*(\mathbf{k}_j) \hat{K}_G(\mathbf{k}_j) e^{-ik_j \cdot \mathbf{q}_i} \\ &= -[K_G * c]^*(\mathbf{q}_i) \\ &= -[K_G * c](\mathbf{q}_i), \end{aligned} \quad (\text{G22})$$

where the convenience function $c(\mathbf{q})$ is defined as

$$c(\mathbf{q}) \equiv h^*(\mathbf{q}) = h(\mathbf{q}), \quad (\text{G23})$$

and the conjugation has no effect because $h(\mathbf{q})$ is real (and $K_G * c$ is as well).

G3.2 Term 2

This term itself consists of several terms, coming from the terms in the sum of equation (G6). In particular, the convolution part $[\textcircled{*}]$,

in full, reads

$$\begin{aligned} [\odot] &= \frac{D_2}{N} \sum_j e^{ik_j \cdot q_m} \hat{K}_G(\mathbf{k}_j) \sum_k e^{-ik_j \cdot q_k} \\ &\times \sum_{a>b} \left(\phi_{,aa}^{(1)}(\mathbf{q}_k) \frac{\partial \phi_{,bb}^{(1)}(\mathbf{q}_k)}{\partial \delta(\mathbf{q}_i)} + \phi_{,bb}^{(1)}(\mathbf{q}_k) \frac{\partial \phi_{,aa}^{(1)}(\mathbf{q}_k)}{\partial \delta(\mathbf{q}_i)} \right. \\ &\quad \left. - 2\phi_{,ab}^{(1)}(\mathbf{q}_k) \frac{\partial \phi_{,ab}^{(1)}(\mathbf{q}_k)}{\partial \delta(\mathbf{q}_i)} \right). \end{aligned} \quad (\text{G24})$$

Like for equation (G11), we can write the three different parts (as defined by the three summands between parentheses) of this equation as four-index quantities:

$$\begin{aligned} \frac{D_2}{D_1} \eta_{cd}^{ab}(\mathbf{q}_i) &\equiv \sum_m h(\mathbf{q}_m) \frac{D_2}{N} \sum_j e^{ik_j \cdot q_m} \hat{K}_G(\mathbf{k}_j) \\ &\times \sum_k e^{-ik_j \cdot q_k} \phi_{,cd}^{(1)}(\mathbf{q}_k) \frac{\partial \phi_{,ab}^{(1)}(\mathbf{q}_k)}{\partial \delta(\mathbf{q}_i)} \\ &= \sum_m h(\mathbf{q}_m) \frac{D_2}{D_1} \frac{1}{N^2} \sum_j e^{ik_j \cdot q_m} \hat{K}_G(\mathbf{k}_j) \\ &\times \sum_k e^{-ik_j \cdot q_k} \phi_{,cd}^{(1)}(\mathbf{q}_k) \sum_l e^{ik_l \cdot q_k} \frac{k_{la} k_{lb}}{k_l^2} e^{-ik_l \cdot q_i} \\ &\equiv \frac{D_2}{D_1} \frac{1}{N^2} \sum_j \hat{c}^*(\mathbf{k}_j) \hat{K}_G(\mathbf{k}_j) \sum_k e^{-ik_j \cdot q_k} \phi_{,cd}^{(1)}(\mathbf{q}_k) \\ &\times \sum_l e^{ik_l \cdot q_k} \frac{k_{la} k_{lb}}{k_l^2} e^{-ik_l \cdot q_i} \\ &\equiv \frac{D_2}{D_1} \frac{1}{N} \sum_k b^*(\mathbf{q}_k) \phi_{,cd}^{(1)}(\mathbf{q}_k) \\ &\times \sum_l e^{ik_l \cdot q_k} \frac{k_{la} k_{lb}}{k_l^2} e^{-ik_l \cdot q_i} \\ &\equiv \frac{D_2}{D_1} \frac{1}{N} \sum_l \hat{a}_{cd}^*(\mathbf{k}_l) \frac{k_{la} k_{lb}}{k_l^2} e^{-ik_l \cdot q_i} \\ &\equiv \frac{D_2}{D_1} \eta_{cd}^{*ab}(\mathbf{q}_i) \\ &= \frac{D_2}{D_1} \eta_{cd}^{ab}(\mathbf{q}_i), \end{aligned} \quad (\text{G25})$$

where the conjugation trivially drops out and we further define convenience functions $b(\mathbf{q})$, $a_{cd}(\mathbf{q})$, and $\eta_{cd}^{ab}(\mathbf{k})$ as

$$\begin{aligned} \hat{b}(\mathbf{k}) &\equiv (\hat{c}^*(\mathbf{k}) \hat{K}_G(\mathbf{k}))^* = \hat{h}(\mathbf{k}) \hat{K}_G(\mathbf{k}), \\ &\Rightarrow b(\mathbf{q}) = [K_G * h](\mathbf{q}) \end{aligned} \quad (\text{G26})$$

$$a_{cd}(\mathbf{q}) \equiv (b^*(\mathbf{q}) \phi_{,cd}^{(1)}(\mathbf{q}))^* = b(\mathbf{q}) \phi_{,cd}^{(1)}(\mathbf{q}), \quad (\text{G27})$$

$$\begin{aligned} \eta_{cd}^{ab}(\mathbf{k}) &\equiv \left(\frac{k_a k_b}{k^2} \hat{a}_{cd}^*(\mathbf{k}) \right)^* = \frac{k_m k_n}{k^2} \hat{a}_{cd}(\mathbf{k}) \\ &\Rightarrow \eta_{cd}^{ab}(\mathbf{q}) = \partial_a \partial_b \nabla^{-2} \left([K_G * h](\mathbf{q}) \phi_{,cd}^{(1)}(\mathbf{q}) \right). \end{aligned} \quad (\text{G28})$$

Here, we can again easily get rid of the conjugations. Apart from obvious term-wise double conjugations, the first term contains the Gaussian kernel, which is real, $\phi_{,cd}^{(1)}$ is also real, as it is a physical term, and the \mathbf{k} terms are real scale vectors. Finally, we replace η by χ :

$$\eta_{cd}^{ab}(\mathbf{q}) = \chi_{cd}^{ab}(b(\mathbf{q}); \mathbf{q}). \quad (\text{G29})$$

Putting it all together, term ②, in full, reads

$$\textcircled{2} = \frac{D_2}{D_1} \sum_{a>b} \left(\chi_{bb}^{aa} + \chi_{aa}^{bb} - 2\chi_{ab}^{ab} \right) (b(\mathbf{q}_i); \mathbf{q}_i), \quad (\text{G30})$$

which is almost the same as the previous 2LPT part in equation (G15), but with a Gaussian smoothed version of h , namely b .

G3.3 Term 3

The simplest term, a matter of a sum over an expression with a Kronecker delta that changes the \mathbf{q}_m 's to \mathbf{q}_i 's:

$$\textcircled{3} = \frac{h(\mathbf{q}_i)}{\sqrt{1 - \frac{2}{3}\delta(\mathbf{q}_i)}}. \quad (\text{G31})$$

G3.4 Term 4

$$\begin{aligned} \textcircled{4} &= - \sum_m h(\mathbf{q}_m) \left[K_G * \left(\left(1 - \frac{2}{3}\delta \right)^{-\frac{1}{2}} K_i^K \right) \right] (\mathbf{q}_m) \\ &\equiv - \frac{1}{N} \sum_m h(\mathbf{q}_m) \sum_j e^{ik_j \cdot q_m} \hat{K}_G(\mathbf{k}_j) \\ &\quad \times \sum_k e^{-ik_j \cdot q_k} \delta_{q_k, q_i}^K \Delta(\mathbf{q}_k) \\ &= - \frac{1}{N} \sum_j \hat{c}^*(\mathbf{k}_j) \hat{K}_G(\mathbf{k}_j) \sum_k e^{-ik_j \cdot q_k} \delta_{q_k, q_i}^K \Delta(\mathbf{q}_k) \\ &= - \sum_k [K_G * c]^*(\mathbf{q}_k) \Delta(\mathbf{q}_k) \delta_{q_k, q_i}^K \\ &= - \Delta(\mathbf{q}_i) [K_G * c](\mathbf{q}_i), \end{aligned} \quad (\text{G32})$$

where we used the convenience function $\Delta(\mathbf{q})$

$$\Delta(\mathbf{q}) \equiv \left(1 - \frac{2}{3}\delta(\mathbf{q}) \right)^{-1/2}. \quad (\text{G33})$$

It may, on first hand, seem counterintuitive that while we started with a convolution of the Gaussian with one function ($\Delta(\mathbf{q}) \delta_{q, q_i}^K$), we end up convolving another function ($c(\mathbf{q})$) and do not convolve with the original one at all. This is actually easily understandable when you consider that what we were calculating is a double convolution. In fact, we could derive the term in another way, without going to Fourier space, but instead only using the definition of a discrete convolution:

$$\begin{aligned} \textcircled{4} &= - \sum_m h(\mathbf{q}_m) \left[K_G * \left(\left(1 - \frac{2}{3}\delta(\mathbf{q}_m) \right)^{-\frac{1}{2}} \delta_{q_m, q_i}^K \right) \right] \\ &= - \sum_m h(\mathbf{q}_m) \sum_j K_G(\mathbf{q}_m - \mathbf{q}_j) \Delta(\mathbf{q}_j) \delta_{q_j, q_i}^K \\ &= - \sum_j [K_G * h](\mathbf{q}_j) \Delta(\mathbf{q}_j) \delta_{q_j, q_i}^K \\ &= - [K_G * h](\mathbf{q}_i) \Delta(\mathbf{q}_i), \end{aligned} \quad (\text{G34})$$

where in the third step we used that $K_G(\mathbf{q}) = K_G(-\mathbf{q})$. Term ④ is a convolution of $h(\mathbf{q})$ with the convolution of K_G with the other term, and convolutions are associative. However, because that latter term contains a Kronecker delta, things get a little bit funky, as they always do when delta functions turn up. With this derivation, we also

immediately see that the $c(\mathbf{q})$ function is superfluous, as we already concluded when we first encountered it in term ① (equation G22). Actually, term ① could also easily have been derived with this double convolution method, circumventing the complication of the Fourier transform of the Kronecker delta function.

G3.5 All terms together

Putting it all together, we obtain for *minus* the derivative of the likelihood in a 2LPT + SC model:

$$F_i^{\mathcal{L}, 2\text{LPT}+\text{SC}} = -[K_G * h](\mathbf{q}_i) + \frac{D_2}{D_1} \sum_{\underline{a} > \underline{b}} \left(\chi_{\underline{bb}}^{\underline{aa}} + \chi_{\underline{aa}}^{\underline{bb}} - 2\chi_{\underline{ab}}^{\underline{ab}} \right) (b(\mathbf{q}_i); \mathbf{q}_i) + \left(\frac{h(\mathbf{q}_i) - [K_G * h](\mathbf{q}_i)}{\sqrt{1 - \frac{2}{3}\delta(\mathbf{q}_i)}} \right), \quad (\text{G35})$$

with $\chi_{\underline{cd}}^{\underline{ab}}(f(\mathbf{q}); \mathbf{q})$ in equation (G14) and $b(\mathbf{q})$ in equation (G26). There are no conjugations left and the remaining c functions have been replaced by h .

Note that in the last term, the square root has an upper bound: $\delta(\mathbf{q}_i) \leq \frac{3}{2}$. If it is larger, the square root will become imaginary.

In fact, this is supposed to happen; the lower bound for the stretching parameter \mathfrak{S}^{SC} is -3 (Neyrinck 2013). This means that the given formula is only valid up to the given upper bound. After that, the entire term (i.e. the sum of terms ③ and ④) becomes zero – the derivative of -3 – i.e.

$$F_i^{\mathcal{L}, 2\text{LPT}+\text{SC}} = -[K_G * h](\mathbf{q}_i) + \frac{D_2}{D_1} \sum_{\underline{a} > \underline{b}} \left(\chi_{\underline{bb}}^{\underline{aa}} + \chi_{\underline{aa}}^{\underline{bb}} - 2\chi_{\underline{ab}}^{\underline{ab}} \right) (b(\mathbf{q}_i); \mathbf{q}_i) + \begin{cases} \left(\frac{h(\mathbf{q}_i) - [K_G * h](\mathbf{q}_i)}{\sqrt{1 - \frac{2}{3}\delta(\mathbf{q}_i)}} \right), & \text{if } \delta(\mathbf{q}_i) \leq \frac{3}{2}, \\ 0, & \text{otherwise.} \end{cases} \quad (\text{G36})$$

This paper has been typeset from a \LaTeX file prepared by the author.